# Revisiting Species Sensitivity Distribution : modelling species variability for the protection of communities

Guillaume Kon Kam King

# Thèse de l'Université de Lyon

**Présentée**

**devant l'Université Claude Bernard Lyon 1**

pour l'obtention

du diplôme de doctorat

(arrêté du 7 août 2006)

soutenue publiquement le

29 octobre 2015

par

Guillaume KON KAM KING

---

# Revisiting Species Sensitivity Distribution: modelling species variability for the protection of communities

---

Directrices de thèse: Sandrine CHARLES et Marie Laure DELIGNETTE-MULLER
Soutenue le devant le jury composé de :

| | | |
|---|---|---|
| **Mme.** | **Marie Laure Delignette Muller** ..... | Directrice |
| **M.** | **Frédéric Yves Bois** ..... | Rapporteur |
| **M.** | **Peter Craig** ..... | Rapporteur |
| **M.** | **Sylvain Dolédec** ..... | Examinateur |
| **Mme.** | **Sabine Duquesne** ..... | Examinatrice |

# UNIVERSITE CLAUDE BERNARD - LYON 1

| | |
|---|---|
| **Président de l'Université** | **M. François-Noël GILLY** |
| Vice-président du Conseil d'Administration | M. le Professeur Hamda BEN HADID |
| Vice-président du Conseil des Etudes et de la Vie Universitaire | M. le Professeur Philippe LALLE |
| Vice-président du Conseil Scientifique | M. le Professeur Germain GILLET |
| Directeur Général des Services | M. Alain HELLEU |

## *COMPOSANTES SANTE*

| | |
|---|---|
| Faculté de Médecine Lyon Est – Claude Bernard | Directeur : M. le Professeur J. ETIENNE |
| Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux | Directeur : Mme la Professeure C. BURILLON |
| Faculté d'Odontologie | Directeur : M. le Professeur D. BOURGEOIS |
| Institut des Sciences Pharmaceutiques et Biologiques | Directeur : Mme la Professeure C. VINCIGUERRA |
| Institut des Sciences et Techniques de la Réadaptation | Directeur : M. le Professeur Y. MATILLON |
| Département de formation et Centre de Recherche en Biologie Humaine | Directeur : Mme. la Professeure A-M. SCHOTT |

## *COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE*

| | |
|---|---|
| Faculté des Sciences et Technologies | Directeur : M. F. DE MARCHI |
| Département Biologie | Directeur : M. le Professeur F. FLEURY |
| Département Chimie Biochimie | Directeur : Mme Caroline FELIX |
| Département GEP | Directeur : M. Hassan HAMMOURI |
| Département Informatique | Directeur : M. le Professeur S. AKKOUCHE |
| Département Mathématiques | Directeur : M. le Professeur Georges TOMANOV |
| Département Mécanique | Directeur : M. le Professeur H. BEN HADID |
| Département Physique | Directeur : M. Jean-Claude PLENET |
| UFR Sciences et Techniques des Activités Physiques et Sportives | Directeur : M. Y.VANPOULLE |
| Observatoire des Sciences de l'Univers de Lyon | Directeur : M. B. GUIDERDONI |
| Polytech Lyon | Directeur : M. P. FOURNIER |
| Ecole Supérieure de Chimie Physique Electronique | Directeur : M. G. PIGNAULT |
| Institut Universitaire de Technologie de Lyon 1 | Directeur : M. le Professeur C. VITON |
| Ecole Supérieure du Professorat et de l'Education | Directeur : M. le Professeur A. MOUGNIOTTE |
| Institut de Science Financière et d'Assurances | Directeur : M. N. LEBOISNE |

# Acknowledgments

Je souhaite tout d'abord remercier mes deux directrices de thèse, Sandrine Charles et Marie Laure Delignette-Muller à qui le travail présenté ici doit beaucoup. Merci à Sandrine de m'avoir encouragé à tout tenter et donné toutes ces belles opportunités de conférences et rencontres, ainsi que pour ses relectures toujours méticuleuses. Merci à Marie Laure pour son engagement sans failles, son exigence et son humanité, sa patience et son soutien précieux. J'ai non seulement beaucoup appris à ses côtés, côtoyant une scientifique brillante, mais collaborer avec elle a été un vrai plaisir.

I would also like to thank Peter Craig and Frédéric Bois along with the rest of my jury Sylvain Dolédec and Sabine Duquesne for carefully reading the thesis and suggesting well-thought improvements. Sylvain Dolédec mérite des remerciements supplémentaires pour son attention lors des deux comités de pilotage de thèse et ses commentaires avisés à ces occasions.

Je voudrais ensuite remercier de tout mon coeur Philippe Veber qui a joué un rôle bien particulier tout au long de cette thèse, à la fois comme avis extérieur, comme collaborateur, comme soutien jamais à cours de bonnes idées et surtout comme ami. Philippe Veber et Floriane Debaulieu, si ces trois années à Lyon ont été un plaisir, c'est beaucoup grâce à vous !

I also wish to thank Ben Kefford, interacting with him was always both interesting and heartwarming. Ben's deep understanding and wide knowledge gave breadth and perspective to our work, and I was always pleasantly surprised by his sincerity and kindness. I also benefited from his insights when drafting the ideas for the foreword of this thesis.

Je souhaite également témoigner ma reconnaissance à Christophe Piscart pour m'avoir suivi tout au long de cette thèse, et pour m'avoir permis cette collaboration avec Ben. Je voudrais aussi remercier Floriane Larras pour son aide et pour son enthousiasme lors de notre collaboration. Je remercie également Bernard Montuelle, Agnès Bouchez et Frédéric Rimet ainsi que les autres membres de l'INRA de Thonon-les-bains pour leur accueil lors de ma visite et pour l'attention qu'ils ont porté à mes projets.

Je remercie Jeanne Garric avec qui j'ai eu la chance de pouvoir travailler, pour le temps et l'attention qu'elle m'a consacré.

Je voudrais remercier les membres actuels ou passés de l'équipe MEPS avec qui il a été un plaisir de partager ces trois années, Christelle Lopes, Dominique Lamonica, Philippe Ruiz, Carole Forfait, Hugo Terrisse, Kévin Elie-dit-Cosaque, Jean-Christophe

Attard ainsi que Cécile Poix, qui sans être membre de l'équipe à proprement parler, y a néanmoins beaucoup apporté. Un merci également à Elise Billoir qui a été présente régulièrement pendant la thèse pour donner divers conseils, notamment pendant les comités de pilotage.

Ensuite, je voudrais remercier chaleureusement mes collègues du LBBE qui m'ont témoigné une franche amitié, une grande gentillesse, et que pour beaucoup j'aurais aimé avoir le temps de mieux connaître. Parmi eux, je pense à Fanny Pouyet, Magali Semeria, Héloïse Philippon, Thomas Bigot, Aline Muyle, Adil El Fitri, Simon Penel, Bastien Boussau, Rémi Planel, Marc Bailly-Béchet, Nicolas Lartillot, Marie Cariou, Michel Lecoq, Dominique Guyot, Jennifer Morinay, Wandrille Duchemin, Ghislain Durif, Stéphane Delmotte, Morgane Tidière, Marion Roussel, Gabriel Marais, Alice Julien-Laferrière, Aurélie Siberchicot, Murray Patterson, Amandine Fournier, Frédéric Jauffrit, Christophe Plantamp, Clément Goubert, Hélène Lopez-Maestre, Florence Bouheddi, Laurent Jacob, Vincent Miele et tant d'autres. Mes excuses à ceux que j'oublierais de citer ici. Je souhaite également remercier Nathalie Arbasetti et Dominique Mouchiroud pour leur aide à de nombreuses occasions, qui m'a permis de mener ma thèse à son terme dans les meilleures conditions.

I switch back to English to thank Igor Prünster and the statistics group of Collegio Carlo Alberto for welcoming me during the writing period of my thesis. I have been given the means to work in a wonderful place and I have really enjoyed the time spent there. Julyan Arbel and Marco Battiston deserve special thanks, I have really enjoyed our numerous discussions and they have been source of inspiration for my work. I also thank Matteo Morini for introducing me to Igor, and more generally for his wonderful kindness.

J'aimerais également remercier Thomas Koffel et ma soeur Juliette pour leur aide sur l'avant-propos de la thèse.

Enfin, je remercie ma famille pour leur soutien, leur amour et leur attention constante, ainsi que Lætitia à qui je ne peux penser sans commencer à sourire, je n'ai jamais autant ri que depuis que je te connais.

Cette thèse a bénéficié du soutien financier de la région Rhône-Alpes.

# Contents

# Foreword

Biodiversity has proven to be interwoven with ecosystem functioning: plant diversity appear to yield higher productivity than monoculture[Tilman et al., 2001], the presence of species that respond differently to perturbations can stabilize ecosystems[Hooper et al., 2005], species loss can have strong ecosystem effects[Srivastava and Vellend, 2005] and invasion by a single foreign species might alter existing dynamics with either positive or negative effects[Davis, 2011]. Broadly speaking, biodiversity is an essential component of our environment. The signature in June 1992 of the Convention of Biological Diversity [Cropper, 1993] by 168 countries followed by 13 Conferences Of the Parties (COP) organized so far attests the need to protect this environment from destruction by human-related activities. The motivations for preserving it stem either from an utilitarian reasoning to preserve and foster human well-being, including food, security, health and cultural aspirations, or from ethical principles about our place in the world and in evolution. These ethical principles are sometimes presented as a component of human well-being [Watson et al., 2003], since a desire to preserve the world we live in can be said to result from a type of moral and cultural values[1]. Human well-being is commonly invoked as an overarching concept to justify compromises between a reasoned impact on the environment to allow equal development for all and the preservation of ecosystems. This opens the way to admitting that it is not possible to erase completely the influence of human activity on the environment and that the acknowledged target of environmental protection is not the preservation of a pristine environment but the protection of certain identified features which will not be perturbed by reasoned human activity. The selection of these features and determination of the degree of protection they need motivates the attribution of a value depending on their properties. An utilitarian and non-utilitarian value can be defined. The utilitarian value corresponds to some useful service the ecosystems provide, such as food production, decontamination, health benefits or $CO_2$ capture. Those are gathered under the concept of ecosystem services, which encompasses all the benefits people obtain from ecosystems [Watson et al., 2003]. Their utilitarian value can be defined by the market price of an equivalent service performed by alternative means. Ecosystem properties also justify the definition of a non-utilitarian value, pertaining to their importance from

[1]However, this is putting those ethical motives for environmental protection on a par with other cultural values such as potential traditions to exploit environmental resources without regard to sustainability (e.g. excessive whaling, or slash-and-burn farming for communities too large).

a aesthetic, scientific, social, cultural or religious point of view. This non-utilitarian paradigm attributes an intrinsic value to species, landscapes, ecosystem which is unrelated to the potential profit obtained by destroying them. In this case, attributing a price or defining an equivalent value might not be possible, although attempts are proposed building on the social and legal sanctions for deteriorating some feature of the environment [Watson et al., 2003]. The difficulty to quantify the intrinsic value means that this non-utilitarian paradigm does not easily lend itself to cost-effectiveness analyses and calculations for prioritisation. However, the existence of an intrinsic value is largely accepted among the advocates of environmental protection, as can be found in the first line of the preamble of the Convention of Biological Diversity:

> Conscious of the intrinsic value of biological diversity and of the ecological, genetic, social, economic, scientific, educational, cultural, recreational and aesthetic values of biological diversity and its components[. . . ]

It could be argued that intrinsic value also fills the space left by incomplete or imperfect information: the connection between the state of an ecosystem and the services it provides is still poorly understood and thinking that we might be able to protect selectively ecosystem services might be fanciful. Moreover, there might be services still not discovered (examples include new drugs or development of biomimetic technologies) which could be lost in a approach focusing on ecosystem services. On the other hand, it is often assumed that adequately preserving the intrinsic value of an ecosystem will also protect every service we receive from it. This justifies using intrinsic value as the target of protection within a precautionary principle approach.

Once the targets of protection are identified, there are various considerations necessary for implementing this protection. It is recognised that it might not be possible to avoid influencing the environment altogether as agriculture, transportation and exploitation of natural resources are bound to have some effect. An approach commonly adopted is to try to quantify the total effect to the environment and keep it to a reasonable/sustainable level. Furthermore, ethical considerations about human well-being compel us to weight the benefits of protecting the environment against the benefits to populations resulting from exploitation. Choosing to preserve biodiversity for its intrinsic value might clash with political aims of reducing poverty, fostering development, or providing equal standards of living for all. An instance of these dilemmas is the replacement of rainforests by palm oil plantations, which has dire environmental consequences but can benefit the local industry of very poor areas[2]. These considerations make environmental protection a difficult task, needing a variety of tools and methods to assess the both the risks to the environment and effects on human well-being in order to make informed political and management decisions.

Species Sensitivity Distribution (SSD) is one of those tools, whose aim is to protect biodiversity in the most straightforward sense: it is used to assess the effect of any con-

[2]However this conception is debated as it has been noted that the benefits tend to be monopolised by a wealthy minority whereas the majority of the population only suffers from the environmental consequences (no trickle down effect)[Watson et al., 2003].

taminant or stressor to communities[3] of species, its final aim being to find concentrations levels which will preserve the majority of all species. As such, it fits in the framework of protecting the structure of ecosystems for their intrinsic value and ought to be complemented with other tools considering ecosystem-level functioning and human well-being. In its most common form, SSD acknowledges the impossibility to find a concentration which has absolutely no effect. It takes a blind bet that ecosystems have some degree of resilience and that a small perturbation may leave the global equilibrium untouched. SSD assumes a static viewpoint in that it does not account for Darwinian evolution of the species (over several generations some species might become more tolerant to contaminants) nor does it consider the possibility of evolution at the ecosystem level, where some species might come to take other's place because the presence of contaminants has switched the dominance structure. As these phenomenon are not completely understood and difficult to predict, SSD is a crude but useful tool that assesses the effects of a stressor on some selected species. By doing so, it constitutes a first step at measuring the impact of human activity on biodiversity.

SSD is subject to strong practical constraints on the length, the cost of the experiments and the number of animals tested in a context of reduction of animal testing. These constraints motivate efforts focused on a better use of available data. This thesis will present several proposals in that direction. It is written as a data-driven progression across increasingly detailed types of data: point-summarized data, end-of-experiment data and time-resolved data. This actually follows a movement of peeling of layers of summarisation to finish with the raw original data. As the data increases in complexity, we propose a new modelling technique so they are included at best in SSD. With each layer of summarisation removed, we show how to make use of the information discarded to extend the predictive power of SSD.

[3] A community is an assemblage of populations of species occupying the same geographical area and in a particular time; it involves interactions within (intraspecific interactions) and between (interspecific interactions) species in communities, including the distribution, structure, abundance, demography.

Chapter 1

# Introduction to Species Sensitivity Distribution

## 1.1 Species Sensitivity Distribution in risk assessment

The framework for ecological risk assessment, which describes mandatory tests for the commercialisation and use of new chemical substances, consists in two parts which are described in the *guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters* [Aagaard et al., 2013]:

- Assessing exposure to the contaminant, ie. determining time-dependent concentrations in the different compartments of the environment.
- Assessing the effect of these time-dependent concentrations on populations and ecosystems.

Species Sensitivity Distribution is a tool which belongs to the second part. It is frequently used as an intermediate stage of a tiered approach for ecological risk assessment.

### 1.1.1 Tiered risk assessment

The tiered structure for risk assessment (Figure 1.1) consists in increasingly detailed appreciations of exposure and effect which all have the same protection goal, and result in the determination of a safe concentration for the environment. The tiers rank with increasing precision, cost and complexity, and with decreasing safety margins (conservativeness) so that in principle, safely passing the first tier guarantees passing all the superior tiers. This structure is intended to provide a cost effective procedure for an adequate environmental protection.

The 1$^{st}$ tier consists in selecting the most sensitive species from standard laboratory species (core toxicity data) and applying an ad-hoc assessment factor. For the 2$^{nd}$ tier, additional data can be collected to form a sample representative of the community of species to be protected and an extrapolation is performed to represent the sensitivities of all species in the community. This extrapolation can consist in taking the geometric mean

of all species sensitivities, or in modelling explicitly the species sensitivity distribution in the community. The 3rd tier relies on artificially constructed model ecosystems which contain an assemblage of species representing several trophic levels and can simulate environmentally realistic exposure regimes [Aagaard et al., 2013]. The 4th tier represents a combination of approaches to refine the degree of realism either in terms of exposure or ecological relevance of the community of species [Brown et al., 2009].

A particularity of the hierarchical tiered structure is that the procedures for each tiers are more and more loosely defined. While minimum data requirements are specified for the second tier, there is some degree of latitude on the methodological choices for SSD. For the higher tiers, only general guidelines are provided, along with a recommendation for maximum transparency.

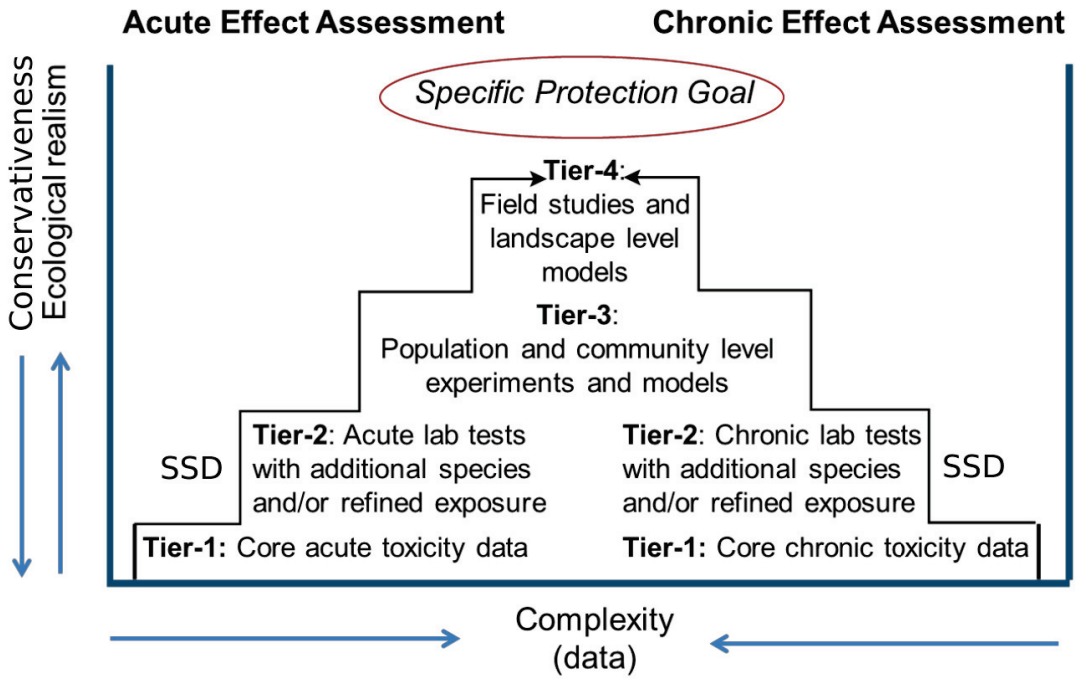



Figure 1.1: Four tiers of ecological risk assessment. Complexity, need for data and ecological realism increase with the tier level (up and towards the centre of the figure), while conservativeness decreases. Figure adapted from [Aagaard et al., 2013]

### 1.1.2 Toxicity data

The first two tiers use so-called toxicity data, which are produced during laboratory experiments. Performing a laboratory experiment to characterise the effect of a contaminant for one species (the precise word is bioassay) means selecting how to measure an effect (choosing an endpoint): the contaminant can affect the survival of a species, its reproduction, growth, behaviour or any physical property. The other decisions to be made regard the concentration range, the duration of the experiment, the

temperature and all the experimental conditions which are selected to maximise the chance of observing an effect. Bioassays to produce toxicity data result in concentration - response or concentration - effect curves, which then contain all the available information on the species response to the contaminant[1]. These curves are often summarised by a single value which will act as a surrogate for the tolerance of the species and might be stored in toxicity databases [RIVM, 2005, eco, ]. These single values are concentrations of interest, a concentration producing a given level of effect (Effective Concentration at $x\%$ ($EC_x$)) or a concentration without effect (No Effect Concentration (NEC)). A practical term to refer to these concentrations is Critical Effect Concentration (CEC)[Forfait-Dubuc et al., 2012]. Another common type of CEC is the No Observed Effect Concentration (NOEC). The concept of NOEC has been disparaged extensively during the last decades because they are based on a wrong interpretation of statistical tests (no statistically significant effect does not mean no effect), they are strongly dependent on the experimental setting and they favour poor resolution on the concentration scale [Chapman et al., 1996, Warne and Van Dam, 2008, Delignette-Muller et al., 2011, Fox et al., 2012, Fox, 2008, van der Hoeven, 1997]. Nonetheless, they do linger in regulatory literature and are worth mentioning for that reason. CECs can be estimated by fitting a model to the concentration - effect curve and calculating the CEC as a function of the model parameters. Common models include the four-parameter log-logistic model [Ritz, 2010]:

$$f(x) = \frac{d-c}{1+(x/e)^b} + c \tag{1.1}$$

where $f$ is the measured endpoint, $x$ the concentration, $c$ the asymptotic value of the endpoint when the concentration grows to infinity, $d$ the value of the endpoint at zero concentration, $e$ the Effect Concentration at 50% ($EC_{50}$) and $b$ a shape parameter,

and the Pires-Fox model [Pires et al., 2002, Fox, 2010]:

$$f(x) = f_0 e^{-b(x-\mathrm{NEC})_+} \tag{1.2}$$

where $f$ is the measured endpoint, $x$ the concentration, $f_0$ the value of the endpoint at zero concentration, NEC is the No Effect Concentration, $(x)_+$ is a shorthand for the maximum between 0 and $x$ and $b$ is a shape parameter.

As estimates obtained from a model fit, these CECs are uncertain and come with a confidence interval.

### 1.1.3 What is the Species Sensitivity Distribution

Due to its position near the base of the tiered system, SSD is routinely used in risk assessment and thus represents an important topic of research. The principle of a species sensitivity distribution approach to risk assessment is to select a few species to build a representative sample of a community to protect, to carry out bioassay experiments for

[1]The distinction between response and effect concerns the nature of the measured endpoint: response designate a quantal outcome (typically survival or death) whereas effect refers to a graded outcome (typically length, mass, growth etc.).

each of them in order to estimate their tolerance and to fit a distribution to describe the whole community. From this distribution, a safe concentration can be estimated, often the Hazardous Concentration for 5% of the species ($HC_5$), the concentration at which 95% of the species should not be affected by the concentration.

The tolerance for each species is obtained by selecting an endpoint and estimating a CEC. When it is not possible to fit a model to estimate the CEC, it is often possible to determine bounds on the CEC. With a collection of CEC for a representative sample of species, the next step is to fit a distribution to extrapolate the sensitivity of the community of species from which the sample originates. Several options are available for selecting the distribution, parametric or distribution-free. The various methods in use to perform an SSD are reviewed in section 1.2.

For illustrative purposes, we present a very simple example of the classic SSD approach. The SSD method starts from laboratory experiments. In this example we present a toy dataset, where 10 hypothetical animals were exposed to increasing concentrations of a contaminant. Survival was measured after a period of time, and the number of survivors was recorded. The data was then presented as concentration-response curves (Figure 1.2).

To estimate a CEC for each species, a three-parameter log-logistic model (Equation 1.1 with $c = 0$) was fitted by maximum likelihood using the R package *drc* [Ritz and Streibig, 2005] (Figure 1.3). The CEC chosen was the Lethal Concentration for 50% of the organisms ($LC_{50}$), the concentration inducing a 50% reduction in survival probability. The last step of SSD is to fit a distribution, often log-normal, to the CEC that were estimated. On this example, the SSD was fitted by maximum likelihood (see section 1.2.3) using the R package *fitdistrplus* [Delignette-Muller and Dutang, 2015]. Usually, the uncertainty on the CEC is not taken into account and the CEC are considered as exact values. The $5^{\text{th}}$ percentile of the distribution is computed to define the $HC_5$ (Figure 1.4). The confidence interval around this $HC_5$ can be estimated in a number of ways. Wagner and Lokke[Wagner and Lokke, 1991] provided an analytical formula for the log-normal distribution, Aldenberg and Slob tabulated values for the log-logistic distribution and bootstrap can be used to approximate the confidence interval for any distribution provided the number of species is large enough[Kon Kam King et al., 2014]. Wheeler[Wheeler et al., 2002] mentions the possibility to use the lower bound of the confidence interval on the $HC_5$ instead of the $HC_5$ as the basis for environmental protection.

### 1.1.4 How Species Sensitivity Distribution is used

Nico van Straalen [van Straalen, 2010] defines a *forward* and *inverse* approach to SSD (Figure 1.5). The *forward* approach consists in predicting the risk of a community subjected to a known concentration in contaminant, based on the SSD. As SSD does not take time into account, this is restricted to constant concentration scenarios.

If $f$ is the probability density of the CEC in the community, for a constant concentration $c$ present in the environment, Van Straalen defines a risk $\delta_c$ as:

$$\delta_c = \int_0^c f(x)dx \tag{1.3}$$

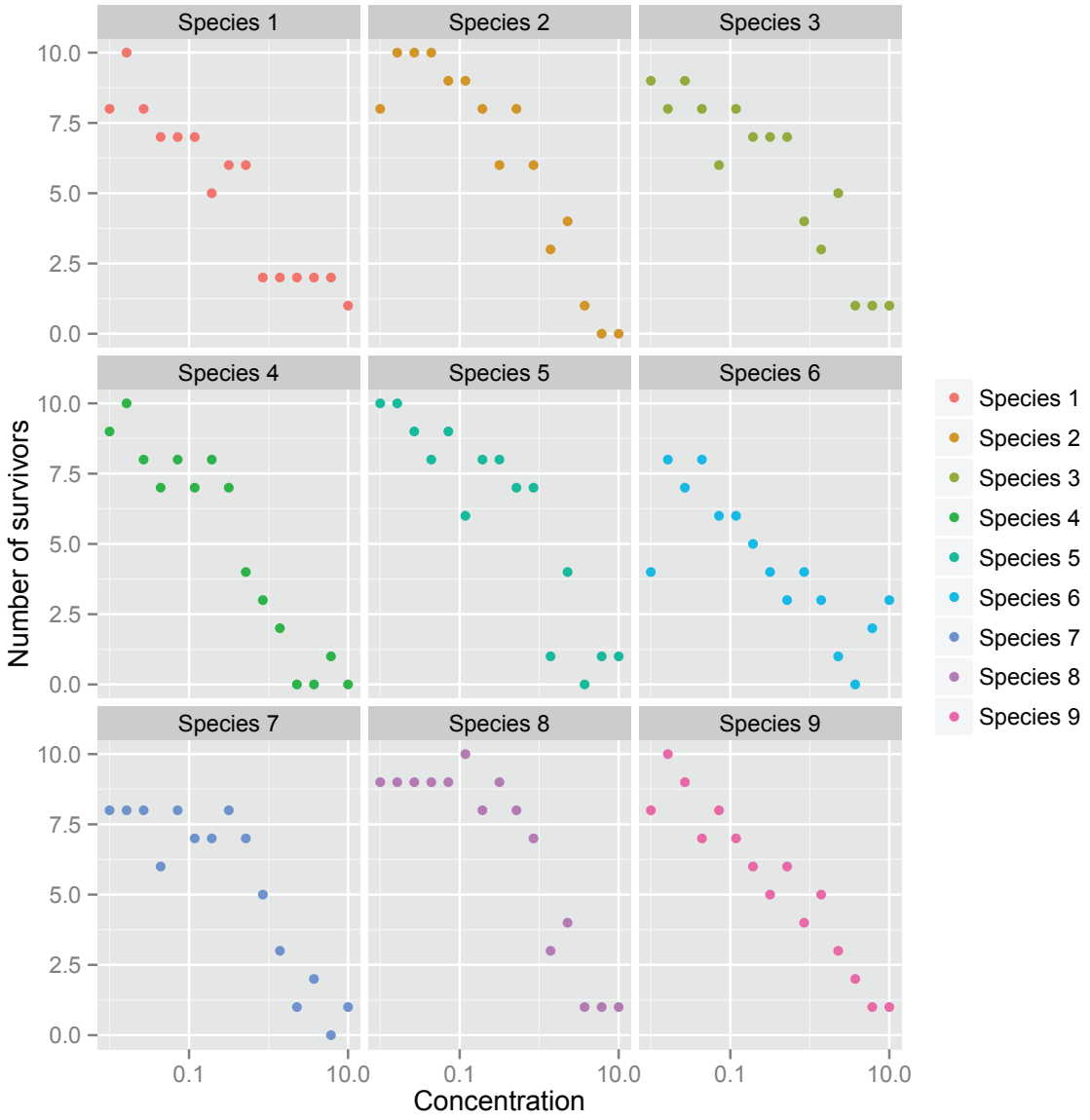



Figure 1.2: Survival as a function of the concentration in contaminant for 9 simulated species.

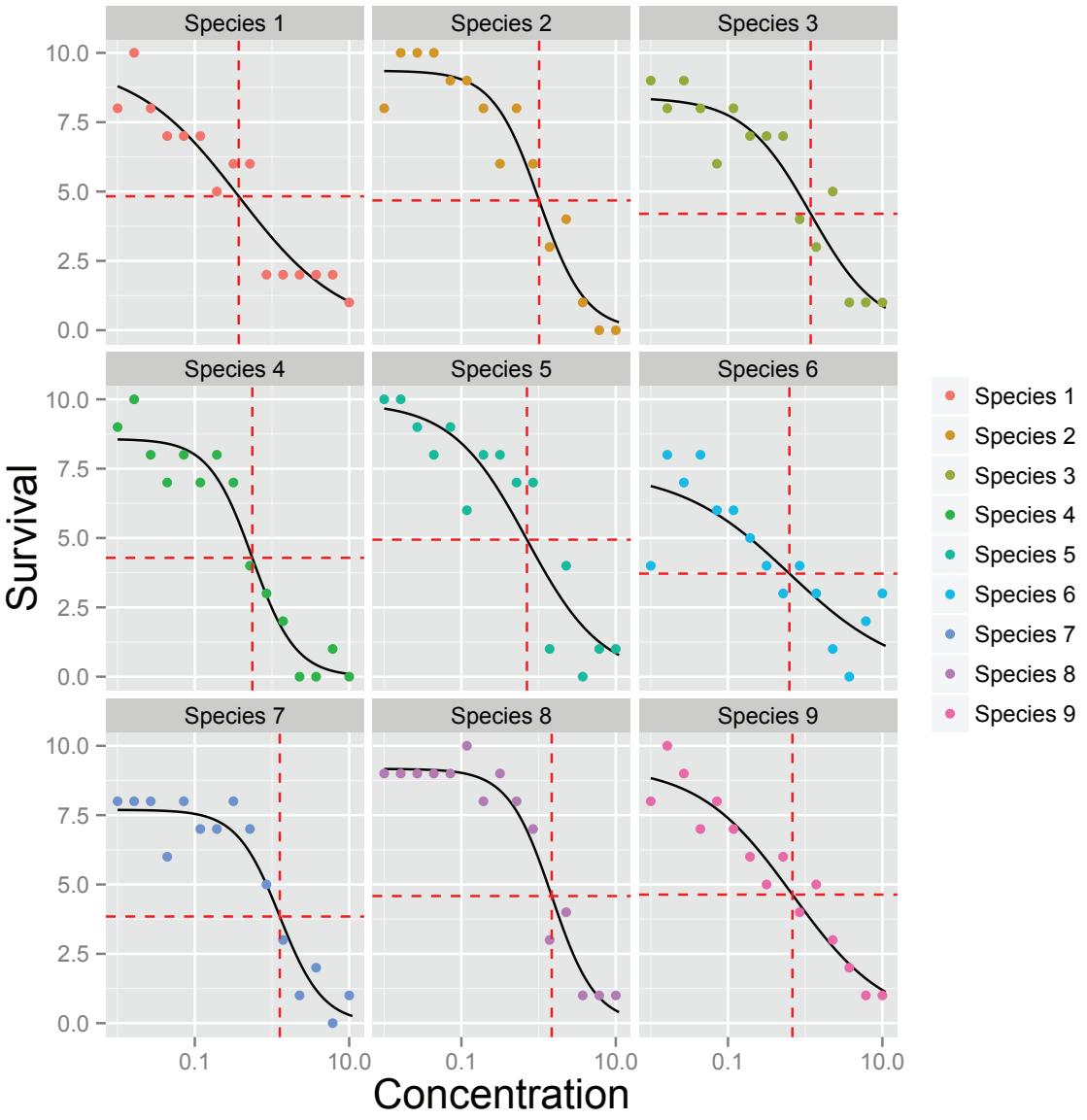



Figure 1.3: Data and fit of a three-parameter log-logistic model for each species. Horizontal red dotted lines correspond to a survival reduction of 50% compared to the control experiment. Vertical red dotted lines correspond to the $LC_{50}$, the concentration for which survival is reduced by 50%.

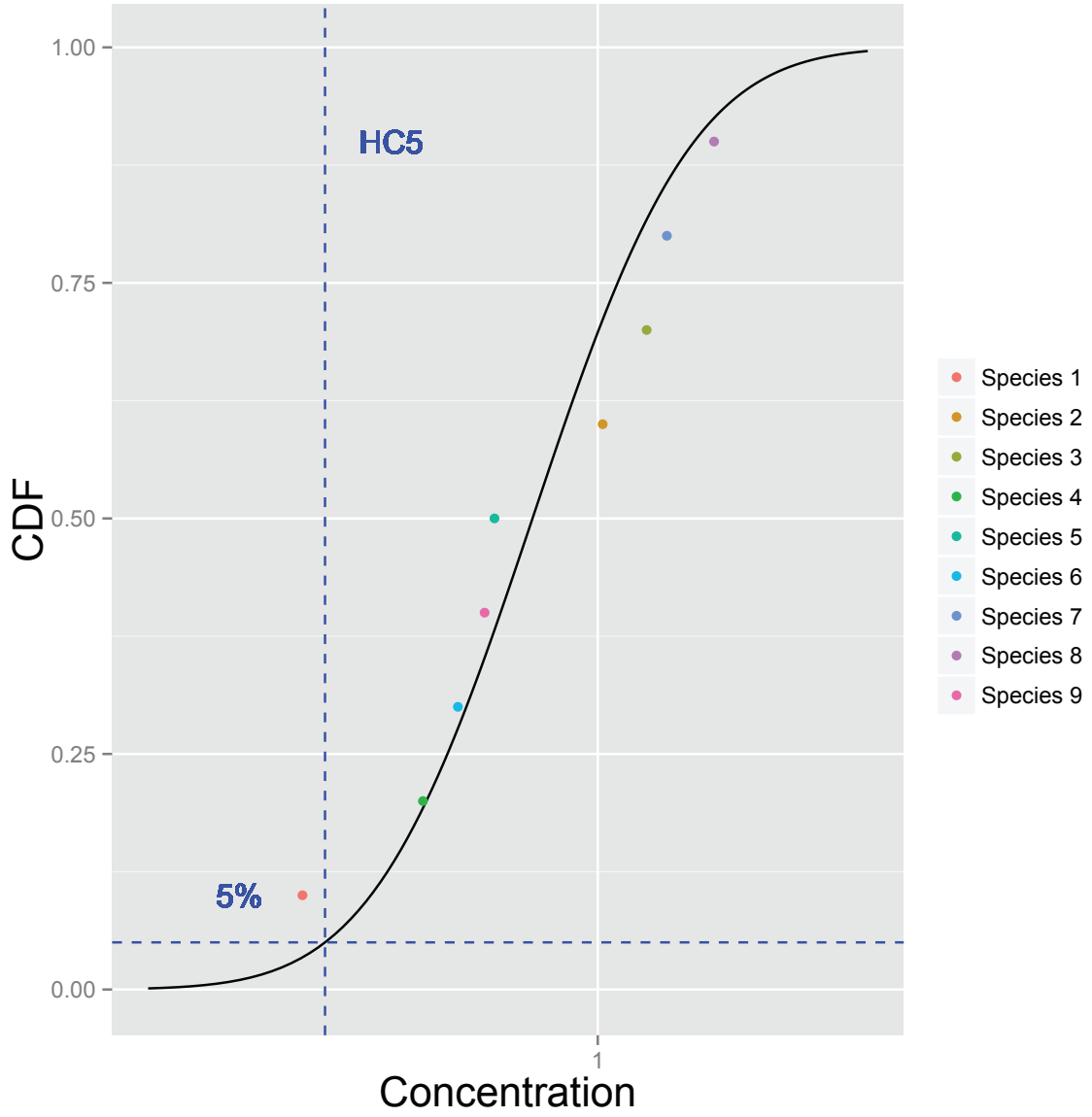



Figure 1.4: Fit of a lognormal distribution on the CEC, with the $LC_{50}$ for each species. The distribution is represented as the Cumulative Distribution Function (CDF).

which can be understood as the proportion of species whose CEC is below concentration $c$. This is also termed affected fraction, or Potentially Affected Fraction (PAF) [Posthuma et al., 2010] to stress the probabilistic nature of this risk assessment.

The second use of SSD, the *inverse* approach, aims at defining an admissible concentration for an arbitrary level of risk which is deemed acceptable. Leaving 5% of the community at risk is the customary choice, based on the hypothesis of a certain redundancy in the role of the species of the community[Forbes and Calow, 2002]. The corresponding admissible concentration is the $HC_5$, the concentration protecting 95% of the community. Note that the meaning of the term "at risk" depends on the CEC chosen. For instance with the $LC_{50}$, leaving 5% of the community at risk means that 5% species might have less than 50% survival, but that may be anywhere between 50% or 100% which does not correspond necessarily to a full elimination of the species. Moreover, the $HC_5$ are used in combination with an assessment factor to account provide a better protection than the bare $HC_5$.

## 1.2 Different flavours of classical Species Sensitivity Distribution

### 1.2.1 Today's use of Species Sensitivity Distribution

SSD is one of the recommended tools for environmental protection agencies worldwide: Australia and New Zealand [ANZECC, 2000], Canada [CCME, 2007], China [Liu et al., 2014], European Union (EU) [ECHA, 2008], South Africa [Roux et al., 1996], United States of America (US) [USE, 1998] all refer to SSD in their guidance documents. These agencies do not always make precise recommendations on the specifics of the method, except from minimum sample size, but rather insist on traceability and transparency. For instance, European Chemical Agency (ECHA) recommends [ECHA, 2012] to use SSD for risk characterisation of short-term or long-term environmental risk, then further refer to the work of [Aldenberg et al., 2002] and to its implementation in the ETx software [van Vlaardingen et al., 2004] for transparency of the statistical analysis. [Nugegoda and Kibria, 2013] provides a survey of the SSD requirements of several agencies in the world and reports that the minimum number of values ranges from 5 in Australia (in 2001) to 10 in the European Union (in 2003), with various requirements concerning the type of data (a summary table from their survey is reproduced in Appendix A).

It is possible to classify most variants of SSD by:

- the choice of a shape for the SSD
- the method to fit the SSD
- the method to account for external information in the SSD

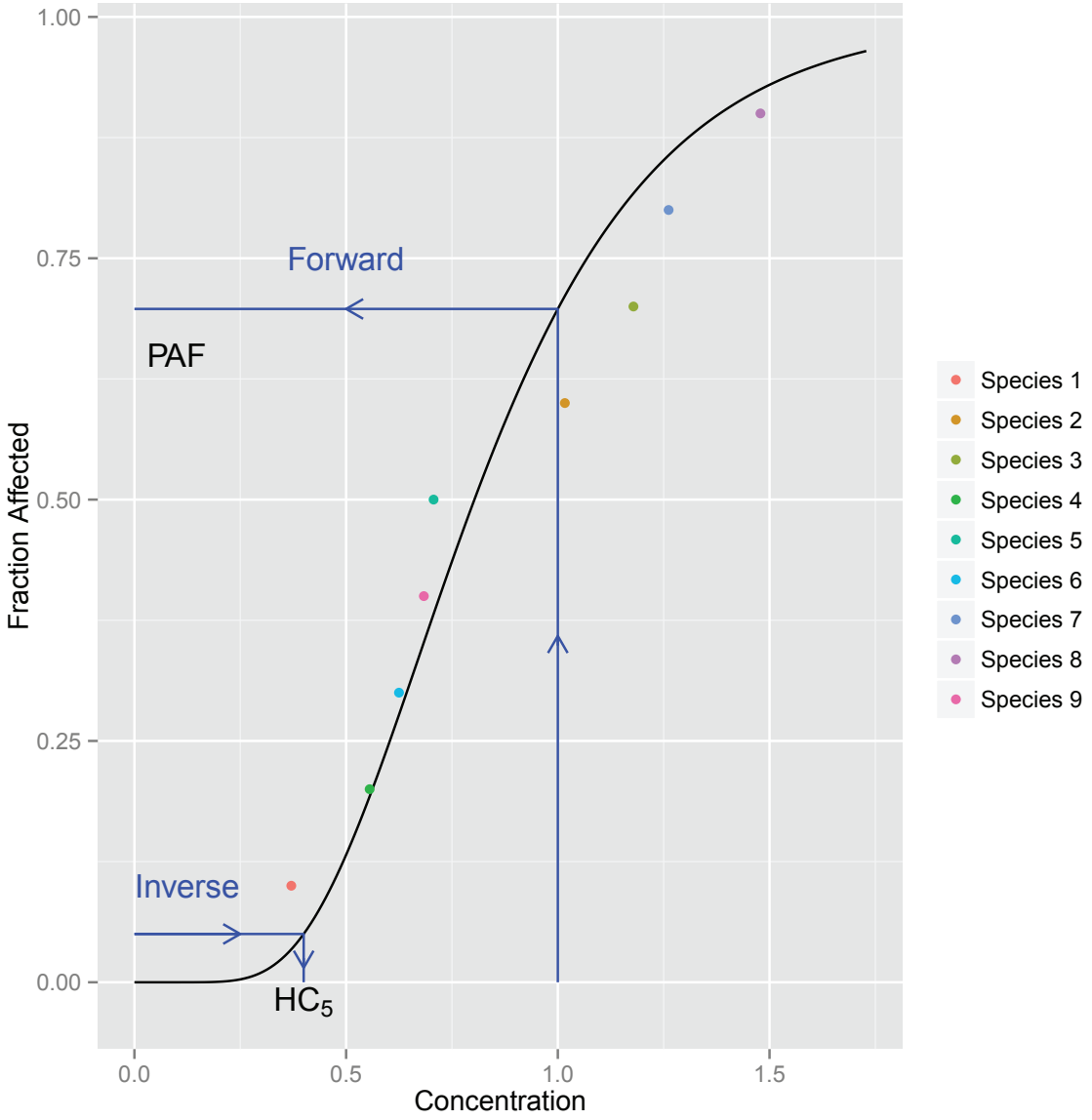



Figure 1.5: Schematic illustration of the *forward* and *inverse* use of SSD, based on the previous example. Figure adapted from [Hickey, 2010].

### 1.2.2 Choice of a shape for the Species Sensitivity Distribution

A first difference among SSD approaches is whether they assume a structure for the sensitivities of the species in the community. One can assume that they are related in that they come from a single parametric distribution, or from a mixture of distributions, but since these assumptions do not rest on any particular ground one may wish to use distribution-free methods instead.

**Different parametric distributions for the Species Sensitivity Distribution**
Several distributions[2] have been proposed for SSD. The most used distribution is the log-normal distribution because it is easy to compute confidence intervals on the $HC_5$. [Wagner and Lokke, 1991] explained that the quantiles of the normal distribution had a non-central t distribution, for which there are analytical confidence intervals. By scaling and centring every distribution, it is possible to tabulate factors and to compute easily the confidence interval for any SSD. Aldenberg and Jaworska [Aldenberg and Jaworska, 2000] showed that from the Bayesian viewpoint, the same confidence intervals could be obtained by choosing a non informative prior on the position and scale parameters of the normal distribution. Aldenberg and Rorije [Aldenberg and Rorije, 2013] later proposed to replace the non informative prior by a uniform prior and derived new factors for calculating the confidence intervals for the log-normal distribution. Note that in their approach, the toxicity data themselves do not follow a normal distribution (they are strictly positive), but log-transformed toxicity data can follow a normal distribution. Equivalently, toxicity data can be said to follow a log-normal distribution and we will use this terminology throughout.

The log-logistic distribution is another widely used distribution (actually the first to be used for SSD in Europe by Kooijman[Kooijman, 1987]), although it does not allow for an analytical expression of the confidence interval on a quantile. Aldenberg and Slob [Aldenberg and Slob, 1993] approximated tables for the one-sided left confidence interval on the $HC_5$ for various sample sizes using a Monte-Carlo procedure. The log-logistic has been advocated over the log-normal distribution because it provides inbuilt conservatism [Aldenberg and Slob, 1993]: having heavier tails, the lower bound of the confidence interval on the $HC_5$ obtained fitting a log-logistic is lower than that obtained fitting a log-normal distribution. Note that the $HC_5$ itself, computed with the [Aldenberg and Slob, 1993] method (a moment-matching method), is bigger though [van der Hoeven, 2001]. This is easier to understand by reasoning on log-transformed data: having heavier tails, the logistic distribution is narrower around its mean than the normal distribution. This implies that below the median, the lowest quantiles of the logistic distribution are smaller than those of the normal distribution while quantiles closer to the median are higher (Figure 1.6). Symmetry of the distribution entails that the situation is reversed above the median. Using numerical root finding, it is possible to determine quantile for which the two distributions coincide: it is approximately the $4^{\text{th}}$ percentile.

Shao [Shao, 2000] proposed to use a distribution with one parameter more than the log-normal and log-logistic distributions, the Burr type III distribution which is commonly

[2]The definition of all the distributions mentioned here can be found on Table 1.1.

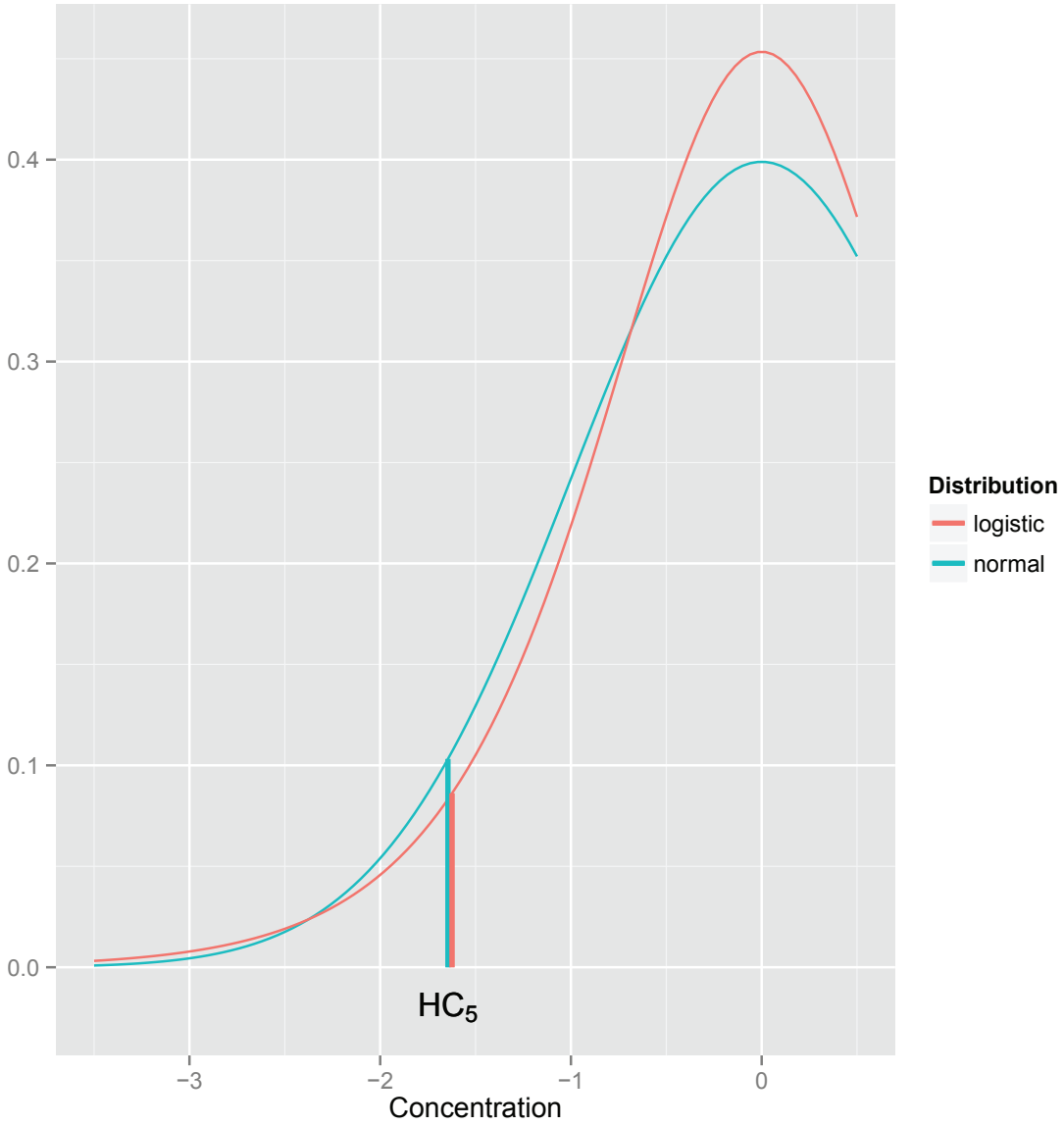



Figure 1.6: Normal and logistic distributions with the same mean and variance and their $5^{\text{th}}$ percentile. The logistic distribution has heavier tails and is also more narrow towards the mean of the distribution.

used in engineering and economy. The common log-logistic distribution is a limiting case of the Burr type III distribution, as are the reciprocal Weibull and reciprocal Pareto. The Burr type III distribution is more flexible than the log-logistic in terms of skewness and kurtosis, but it is also harder to fit due to the extra parameter, particularly in the case of small datasets. To alleviate the difficulties of parameter estimation in the presence of strong structural correlation, [Shao, 2000] proposed to fit a Burr III distribution nonetheless and to then switch to one of the limiting distribution if the values of the estimated Burr III parameters suggest it. Confidence intervals are calculated either using the delta method [Casella and Berger, 2002] or bootstrap in the case of small sample sizes. This distribution and its limiting cases have been included in the Burrlioz software [Campbell et al., 2000] and are recommended for use in Australia and New Zealand.

[van Straalen, 2002] proposed to fit threshold distributions, to remove the need of computing an $HC_5$ or any cut-off point in the distribution. This is interesting for two reasons: 1) choosing a cut-off, for instance 5% of species that will not be protected, is unsatisfying as these species may play an essential role in the ecosystem. Their decrease in abundance might have cascading effects on the other species of the community which might result in a more severe harm than expected with an SSD approach. 2) Infinite tail distributions are not realistic as they allow for infinitely sensitive species for any contaminant. This is certainly false for essential metals that are naturally present in the environment and for which all species in the community must have a degree of tolerance. Among those threshold distributions, Van Straalen tried the log-uniform, log-triangular, log-exponential and log-Weibull distributions. His conclusion was that among the threshold models, the triangular distribution seems to provide the best fit, and that the estimated threshold ($HC_0$) is not very different from the log-normal or log-logistic $HC_5$.

[Zajdlik et al., 2009] proposed to fit a multimodal distribution to the toxicity data. They quoted that when the dataset shows signs of multimodality, the recommended practice is to select a taxonomic subset of the species. However, using a multimodal distribution makes it possible to use all the data available, not only a subset. They proposed selecting the number of components with model comparisons based on Quantile-Quantile plot (Q-Q plot)[3], Kolmogorov-Smirnov tests and likelihood ratios. They developed their approach with a case study on the herbicide atrazine with a dataset containing target and non-target species. The sensitivity of this community showed visual evidence of multimodality which they tested using Hartigan's dip test[4] [Hartigan and Hartigan, 1985]. On their particular dataset, their selection procedure yielded a two component mixture model (on the log-transformed data) which they fitted using likelihood optimisation. They observed that the bimodal distribution provided a better fit than a number of unimodal distributions and that their SSD model predicted an $HC_5$ comparable to the

[3] A Q-Q plot is a plot of the theoretical quantiles of a fitted distribution against the observed quantiles of the sample. The fit of the distribution is assessed by comparing the points to a line representing exact match between the theoretical and observed quantiles.

[4] Hartigan's dip test consist in finding an unimodal distribution function which minimizes Kolmogorov distance between that unimodal distribution function and the empirical distribution function. That minimum distance is the *dip* of the distribution, which measures departure from unimodality and is used as the test statistic.

concentration at which effect start appearing in field experiments. Their conclusion was that using a multimodal distribution allowed performing an SSD on all the species when an unimodal distribution did not fit the data, and that the derived $HC_5$ should be more environmentally relevant than that derived on only a subset of the species.

We presented so far the various distributions proposed for SSD. Several comparisons among these distributions were made, with the BurrIII distribution being found to provide a better fit than the log-normal or log-logistic distributions on a large dataset of 89 species exposed to dichlorodiphenyltrichloroethane (DDT) [Xu et al., 2015]. Comparing BurIII, log-logistic, log-normal, reciprocal Weibull and Weibull models on 32 persistent contaminants, He et al. [He et al., 2014] also found that the BurrIII distribution model generally fitted better, also other models might show comparable fit. Among two-parameter distributions, studying 30 datasets Newman et al. [Newman et al., 2000] found that the Gompertz model fitted better than either the log-normal and log-logistic models. Among the threshold models, Van Straalen [van Straalen, 2002] found the triangular distribution to provide the best fit. [Xing et al., 2014] found that the Weibull and log-triangular distributions provided the best fit based on the Akaike Information Criterion (AIC)[5] criterion and Residual Square Error (RSE). The lack of clear winner and the generally small size of the datasets which precludes very informed distributional choices has led to the adoption of the log-normal distribution as the most commonly used parametric distribution [Wheeler et al., 2002]. [Newman et al., 2000] noted that among 30 tested datasets of NOEC data, half failed the shapiro-wilk [Shapiro and Wilk, 1965] test for log-normality. He goes on to suggest the use of non-parametric methods for SSD.

**Non-parametric methods**

Distribution-free methods are more commonly called non-parametric methods, although very different paradigms coexist under the name non-parametric. The adjective non-parametric can refer to methods which avoid altogether the use of any parameter, or to methods for which the structure of the model is not fixed and adapts to the data. Non parametric tests (Spearman, Kruskal-Wallis, ...) fit in the first category, while kernel density estimation or Bayesian non-parametrics fit in the second.

**Distribution-free methods based on rank** A naive way to compute an $HC_5$ would be to take $n$ species, to rank them by sensitivity and take the $5^{\text{th}}$ percentile. $n$ would have to be large enough for this percentile to be defined. Moreover, the definition of the exact value of a given percentile is arbitrary and a choice would have to be made. Several propositions are available, which are presented and discussed in [Aldenberg et al., 2002].

This naive approach was refined in [van der Hoeven, 2001] by using the asymptotic

[5] For a given statistical model fitted on data, the AIC is defined by $AIC = 2k - 2\ln L$ where $k$ is the number of parameters of the model and $L$ is the value of the likelihood at its maximum. The AIC can be used as a tool for model comparison which includes a measure of the fit of the model and a penalisation for model complexity. Indeed, a model with many parameters will tend to fit the data very well but it might have poor predictive power, a phenomenon known as overfitting. A comparison between model 1 and model 2 is achieved by computing $\Delta_{AIC} = AIC_1 - AIC_2$ where the order of subtraction is chosen to have $\Delta_{AIC} > 0$, the best model being the one with the smaller AIC. The strength of evidence of one model against the other is computed as $l = e^{-\frac{\Delta_{AIC}}{2}}$ [Burnham et al., 2011] and allows for statements such as "the evidence for model 1 is $\frac{1}{l}$ times stronger than for model 2".

Table 1.1: Distributions which have been proposed for the SSD, by alphabetical order. The support for all these distributions is $\mathbb{R}^+$.

| Name | CDF | Constraints |
|---|---|---|
| Burr type III | $\frac{1}{\left(1+(x/\alpha)^{-\beta}\right)^k}$ | $\alpha, \beta, k > 0$ |
| Gompertz | $1 - \exp\left(-\eta\left(e^{bx} - 1\right)\right)$ | $\eta, b > 0$ |
| log-Exponential with threshold | $\begin{cases} 0 & \text{for } x < a \\ 1 - \exp\left(-\frac{\ln x - a}{\lambda}\right) & \text{for } \ln x \geq a \end{cases}$ | $a, \lambda > 0$ |
| log-Gumbel | $\exp(-\exp(-\frac{\ln x - \mu}{\sigma}))$ | $\eta, b > 0$ |
| log-logistic | $\frac{1}{1+(x/\alpha)^{-\beta}}$ | $\alpha, \beta > 0$ |
| log-normal | $\frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{\ln x - \mu}{\sigma\sqrt{2}}\right]$ | $\sigma > 0$ |
| log-triangular | $\begin{cases} 0 & \text{for } \ln x \leq a \\ \frac{(\ln x - a)^2}{(b-a)(c-a)} & \text{for } a < \ln x \leq c \\ 1 - \frac{(b - \ln x)^2}{(b-a)(b-c)} & \text{for } c < \ln x < b \\ 1 & \text{for } b \leq \ln x \end{cases}$ | $a \leq c \leq b$ |
| log-uniform | $\begin{cases} 0 & \text{for } \ln x < a \\ \frac{\ln x - a}{b-a} & \text{for } \ln x \in [a, b) \\ 1 & \text{for } \ln x \geq b \end{cases}$ | $a < b$ |
| log-Weibull with threshold and power 2 | $\begin{cases} 0 & \text{for } \ln x < a \\ 1 - \exp\left(-\left(\frac{\ln x - a}{\lambda}\right)^2\right) & \text{for } \ln x \geq a \end{cases}$ | $a, \lambda > 0$ |
| Reciprocal Pareto | $\begin{cases} \left(\frac{x}{x_0}\right)^\theta & \text{for } x < x_0 \\ 1 & \text{otherwise} \end{cases}$ | $x_0 > 0$ |
| Reciprocal Weibull | $\exp(-\alpha x^{-\beta})$ | $\alpha, \beta > 0$ |

properties of ranks in a sample. This refined method is applicable to find the $HC_5$ from sample size 19 and upwards. The first assumption of the method is that the total number of species in the community is very large. If $n$ species are sampled from a population of $N$ with $N >> 1$ and sorted by increasing tolerance, then the relative rank of a given species among the $N$ species is uniformly distributed between 0 and 1. Therefore order statistics and notably the $5^{\text{th}}$ percentile follow a Beta distribution. It is thus possible to compute the $HC_5$ and its confidence interval but by construction, the expectation of the $HC_5$ will be the CEC of one of the tested species, or the interpolation between two CEC. This presents the inconvenient of leaving some of the species explicitly unprotected, and require huge sample sizes to provide a 95% confidence interval (59 tested species for a one-sided 95% confidence interval). The method was found to give $HC_5$ higher than a log-normal SSD and lower than a log-logistic SSD, while the one-sided 95% confidence interval was consistently larger than those of log-normal and log-logistic SSD [van der Hoeven, 2001], which led van der Hoeven to argue that log-normal and log-logistic SSD were over protective.

[Chen, 2004] proposed another non-parametric method based on the tested species'

rank, which also needs samples of at least 19 species. It used an asymmetric loss function for which one must specify: 1) how costly is over-protection relatively to loss of species and 2) what is the threshold below which over-protection entails a cost. He derived a corresponding risk-minimization estimator for the $HC_5$ and compared it to the van der Hoeven $HC_5$ for several parameters of the loss function, then argued that the van der Hoeven $HC_5$ is generally over-protective. Chen gave examples of parameters for which the estimator was both less conservative and had a lower risk for data coming from log-normal and log-logistic distributions. Once again, the method is only applicable to datasets which are very large on current standards. Moreover, it requires arbitrary choices regarding the risk function which are ultimately political, and can only be left to environmental managers. These choices are difficult to calibrate, inasmuch as the risk of losing a species of intrinsic value is difficult to define[6], whereas resorting to a precautionary principle is easier to explain.

**Another non-parametric approach** A last recent non-parametric approach relies on non-parametric kernel estimation [Wang et al., 2015]. It was demonstrated on very large datasets but not compared to other non-parametric methods. The authors argue that kernel density estimation fits these huge datasets better than parametric distributions because of the presence of outliers. The method uses Gaussian kernel functions.

**Bootstrap methods**

Several boostrap methods have been proposed to improve on the previous parametric and non-parametric approaches. The idea behind bootstrap methods is to approximately reproduce the properties of a population by sampling with replacement in a sample of the population. Several types of bootstrap methods for SSD were proposed. [Jagoe and Newman, 1997] proposed a first non-parametric bootstrap approach consisting of repeatedly estimating the $HC_5$ on samples with the same size as the number of species in the dataset. [Grist and Leung, 2002] proposed an improvement on this method using bias-corrected and accelerated estimators for the confidence interval. These corrections ensure that the error in the coverage of the confidence interval decrease with $\frac{1}{n}$ instead of $\frac{1}{\sqrt{n}}$ for the original method. This non-parametric bootstrap method is restricted to large sample sizes. The bare minimum for estimating an $HC_5$ is 20 species, otherwise the 5th percentile of the empirical distribution is not defined. This is already very large on most standards (only 5 % of the 3448 chemicals in the RIVM database published in [Hickey et al., 2012] have 20 or more species), but attempts to define an optimal size resulted in much larger estimates [Newman et al., 2000, Grist and Leung, 2002]. Grist proposed a bootstrap regression method [Grist and Leung, 2002] for small sample sizes, which is effectively a parametric method: it consists in 1) fitting a parametric SSD on many bootstrap samples, then in 2) estimating the $HC_5$ of the parametric SSD and finally

[6] [Hickey et al., 2009] elaborated on the use of loss functions (but in a parametric setting, using a log-normal distribution) and proposed a linear exponential loss function which penalizes exponentially overestimation of the $HC_5$ and penalizes linearly underestimation. This asymmetric penalization is well suited to a situation where overprotection might have significantly less adverse effects on the long term than underprotection (loss of more than 5% species). However there is no escaping the definition of a relative cost of losing species against putting over-restrictive regulation, which is difficult to reconcile with the intrinsic value paradigm in biodiversity protection.

3) calculating the median then the confidence interval from the bootstrap distribution of the parametric $HC_5$. Wheeler[Wheeler et al., 2002] and Grist [Grist and Leung, 2002] report differences between parametric SSD and bootstrap regression with the same parametric distribution which can be large, while several authors report the contrary [A.M Verdonck et al., 2001, Xing et al., 2014].

[Wang et al., 2008] proposes a modified bootstrap which consists, for $N$ data points, in defining an interval around each data point, in sampling randomly with replacement $N$ intervals then for each interval, sampling a point from a uniform distribution on that interval. This amounts to a noisy version of the bootstrap, where each data point is added random noise. Using that procedure, the probability to sample several times the same value vanishes, and the jaggedness of the empirical cumulative distribution is reduced. The length of the interval can be tuned to vary the intensity of noise around the data. Wang et al. argue that their procedure is more robust than non-parametric bootstrap.

### 1.2.3 Different frequentist methods to fit a parametric Species Sensitivity Distribution

As bootstrap methods seem applicable only in the case of large toxicity datasets, which are not the majority, parametric methods are much more commonly used for SSD. Parameter estimation in a frequentist framework can be performed by optimising a chosen criterion: the likelihood or some goodness-of-fit distance.

**Maximum Likelihood**

The likelihood function gives the probability of observing the data given the parameters. Maximizing the likelihood implies selecting the parameters for which the probability of observing the data is highest. Maximum likelihood is by far the most standard approach to distribution fitting and more generally to model fitting. It is backed with a consequent body of theoretical work ensuring many interesting asymptotic properties [Armitage, Peter and Colton, 2005]: the maximum likelihood estimate converges to the true value of the parameters with increasing sample size (consistency), it is the fastest estimate to converge (efficiency) and the difference with the true value is normally distributed (normality). This last property is useful as it provides a method to compute confidence intervals on the parameters. This approach is used by the Australian software Burrlioz [Campbell et al., 2000], for instance, which fits a Burr III distribution to the toxicity data. Moreover, a natural extension of the likelihood function allows to take into account censored data [Helsel, 2005].

**Least-square regression on the Cumulative Distribution Function and on the Quantile-Quantile plot**

Least-square regression on the empirical CDF is a popular method to estimate the parameters of the SSD. It consists finding the best parameters by minimizing the sum of the squared vertical distances between the CDF of the chosen parametric distribution and the empirical CDF of the data (also known as the Cramer-Von Mises distance [Von Mises, 1964]). Alternatively, least-square regression can be performed on the Q-Q plot. A Q-Q plot is a plot of the quantile of the theoretical distribution against those

of the empirical distribution. If the sample follows the theoretical distribution, then the points would lie on a straight line. Regression on the Q-Q plot consists in minimizing the horizontal or vertical sum of squared distances between a straight line and the theoretical quantiles plotted against observed quantiles.

This is the approach adopted by the software CADDIS_SSD [Shaw-Allen and Suter II, ], from US EPA, which performs a backwards least-square regression on the Q-Q plot with a log-probit function [Hickey and Craig, 2012] (this corresponds to assuming a log-normal distribution for the SSD). The software SSD MASTER [Rodney et al., 2008] uses a least-square regression on the CDF, but tries several other distributions : normal, logistic, Gompertz and Gumbel[7]. [van Straalen, 2002] used the Systat - nonlin module to perform nonlinear regression on the CDF.

However, there is no unique way to build a CDF: several possible plotting positions [Kefford et al., 2012a, Posthuma et al., 2010] all have desirable properties, but none of them represent the data more faithfully than any other. Therefore, the resulting SSD and its predictions depend on purely arbitrary decisions regarding the plotting positions, a fact that undermines its scientific credibility. A number of plotting positions are mentioned in [Aldenberg et al., 2002] and most of them have been proposed following a reasoning on the mean, median or mode of the order statistics (the distribution of the $i^{\text{th}}$ value). The different propositions can be presented as special cases of the following formula:

$$p_i = \frac{i - a}{n + 1 - 2a} \tag{1.4}$$

with $0 \leq a \leq 1$ and where each proposition corresponds to a value for $a$. Values for $a$ proposed by several authors include $a = 0, 0.3, 0.31, 0.3175, 0.375, 0.4, 0.44, 0.5, 1$. The choice of a value for $a$ stems from a decision concerning the mathematical properties of the plotting positions: for $a = 0$, the plotting positions will match the expectation value of order statistics for samples drawn from a uniform distribution (the order statistics follow a beta distribution), whereas for $a = 1$ the plotting positions will match the mode of the order statistics distribution instead. Requirement to match the median of the order statistics distribution do not lead to an exact value for $a$. For example, requirement to match the median for different samples drawn for a log-normal distribution would yield different values for $a$, so it is clear that there is not objectively best choice. Moreover, on an example dataset used in [Aldenberg and Jaworska, 2000] we can see that the value of the $HC_5$ can be affected up to a factor 3 by the choice of the plotting positions (Figure 1.7).

[7]They call it Fisher-Tippet, which is another name for the generalized extreme value (GEV) distribution of which the Gumbel distribution is a special case, but it seems that they are using the Gumbel distribution.

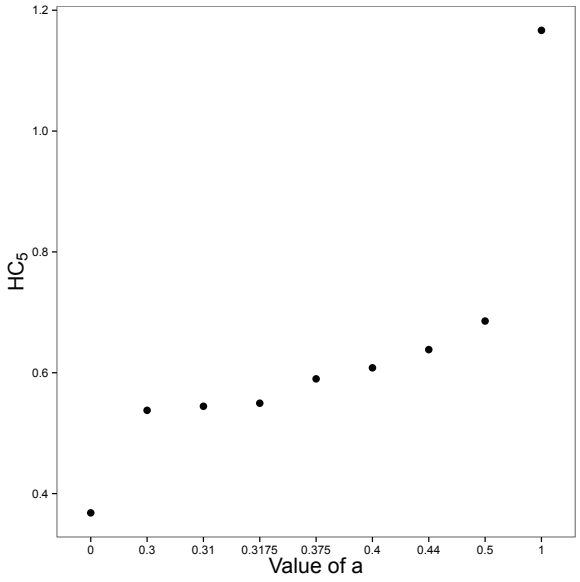



Figure 1.7: HC5 computed on the example cadmium dataset of [Aldenberg and Jaworska, 2000] for several plotting positions, sorted by increasing values of $a$.

Similarly, [A.M Verdonck et al., 2001] shows a difference by a factor 2 between the Hazen plotting positions ($a = 0.5$) and mean plotting positions (Weibull, $a = 0$). Note however that the effect of choosing particular plotting positions is expected to decrease as the size of the sample increases.

When the choice of plotting positions is made, it is possible to measure the least-square distance between the theoretical and the plotted empirical CDF or between the Q-Q plot and a straight line. [Hickey and Craig, 2012] cites several limitations to the regression method: 1) as we have seen, the $HC_5$ depends on plotting positions, 2) the typical regression assumption of errors being independently and identically distributed is false and 3) there is no reason to prefer horizontal least squares to vertical least squares minimization. The reason why the regression assumption fails is two-fold: first, the CDF and the quantile function are non decreasing functions, the value of the empirical CDF and of the empirical quantile function at a given observation can only be equal or higher than at the previous observations which is a clear violation of the independent errors assumption. Second, in the case of regression on the CDF, the normal error model is not true, in particular for the most sensitive and most tolerant species. For these extreme species, the CDF is close to 0 or to 1 and bounded by 0 and 1, so the measurement error cannot be symmetric around the model. Finally, according to Hickey [Hickey and Craig, 2012], CADDIS_SSD [Shaw-Allen and Suter II, ] uses horizontal least squares to perform regression on the CDF. Vertical least squares would be more natural since it amounts to consider that the data is randomly distributed rather than the theoretical quantiles. In summary, regression on the CDF or the Q-Q plot is not particularly recommended [Hickey and Craig, 2012], all the more that simple alternatives exist which are much more reasonable from a statistical point of view.

**Moment matching**

Parameter estimation can also be performed without any distance but by matching properties of the empirical sample with properties of the theoretical distribution. ETx [van Vlaardingen et al., 2004], one of the tools recommended by ECHA for SSD estimates the distribution parameters via moment-matching. Earlier versions of the software performed estimation for log-normal, log-logistic and log-triangular distributions while the current version uses only the log-normal distribution. Moment matching consists in equating the empirical moments of the sample (possibly the centred moments and unbiased estimates) with the moments of the distribution fitted to the data. This is practical in the case where the moments of the distribution have a closed form as a function of the parameters, as it is the case for the log-normal and log-logistic distributions. Moment matching is equivalent to maximum likelihood for distributions of the exponential family, such as the normal distribution. However, distributions such as the Burr III , triangular and log-logistic distributions do not belong to this family. Moreover, moment matching is sensitive to outliers and can give unrealistic results, as the estimators are not guaranteed to be consistent.

### 1.2.4 Bayesian methods to fit a Species Sensitivity Distribution

Finally, various Bayesian methods were proposed to fit an SSD, which share the same paradigm but different aims. A presentation of the Bayesian paradigm is provided in Appendix B. One common idea in these Bayesian approaches is the realisation that classical SSD considers each experiment about a chemical as completely independent from the past and general knowledge in ecotoxicology, and considers each species as a random sample from an unknown community. In reality, there is considerable information already available from the testing of other contaminants or from biological knowledge of the species in the community. The flexibility of the Bayesian framework allows extending the SSD model to take various types of information available about the data into account:

**Species non-exchangeability**

The question of species non-exchangeability has been recently discussed by [Craig et al., 2012]. Species exchangeability is the bayesian counterpart to the assumption of independently identically distributed CEC made in frequentist SSD (see also section B.2). Exchangeability is an assumption that knowing additional information about a species, such as its taxonomic group, would not provide a priori information about its sensitivity. Non-exchangeability of species arise where some commonly tested species have a tendency to be more sensitive or tolerant than others across contaminants, such that it is possible a priori to predict that it should be more sensitive than another species. The trout *Oncorynchus Mykiss* is an example of non-exchangeable species, as it appears to be generally more sensitive than other species to contaminants. [Craig et al., 2012] proposed introducing a hierarchical model with two additional parameters for adjusting the mean and the variance of the sampling distribution of the non-exchangeable species. Adjusting the mean amounts to assuming a general tendency for that species, independent of the contaminant, while adjusting the variance can reduce the variability across contaminants. Craig and colleagues derive new decision rules for

SSD with non-exchangeable species which can be used as easily as those presented by [Aldenberg and Jaworska, 2000].

**Various subgroups**

Another approach is proposed by [Hayashi and Kashiwagi, 2010] who suggest models taking into account taxonomic information about the species and dividing them in three groups: they propose to fit 4 models with a single distribution for all species, a mean for each taxonomic group and common variance, a variance for each taxonomic group and common mean, a variance and a mean for each group. Model selection is carried out using Deviance Information Criterion (DIC)[8] [Spiegelhalter et al., 2002]. When there are several distributions, the global $HC_5$ is obtained by simulating the same number of values from each distribution, or by simulating with weights as suggested by [Duboudin et al., 2004b]. The result of Hayashi and Kashiwagi is that 3 times out of 7, the model with only one distribution is chosen, 3 times out of 7, the model with varying means but common variance is preferred and 1 time out of 7, the model with varying mean and variance is preferred. Other proposals for modelling subgroups based on taxonomic information include [Grist et al., 2006] and [Hickey and Kefford, 2008].

**Inclusion of expert knowledge**

Yet another sort of information available about the data is experts' opinion about the general tolerance of species. [O'Hagan et al., 2005] proposes to estimate the sensitivity of each taxon by including expert judgement about the sensitivity of that taxon and assuming homogeneous within taxon variance for the tested species. In their model, the estimated mean is linearly dependant on expert judgement with a normal error model, and the parameters of a taxon sensitivity distribution (SSD, but at the taxon level) are all fitted in a hierarchical model. This approach is also used in [Grist et al., 2006].

[Ciffroy et al., 2012] presented an informative SSD approach based on prior information on other contaminants to estimate the variance. Their model does not take intra-species variation of the data into account. The prior on the variance is estimated from a pooled variance of all other contaminants. Their model allow for within contaminant variance and inter-contaminant variance. This assumes that variance is similar across contaminants and this assumption is found to hold reasonably well on a dataset of 21 substances. They contend that this assumption is not a very strict one, as prior information will shift the posterior estimation, but in a subtle manner: the extremely small or large variances will be shrunk towards the variance of the bulk of the contaminants. In summary, for large sample size, the parameters are learnt from data and the prior has little influence, whereas for samples of small size who naturally tend to have a small variance, the prior pulls toward larger variance which is a conservative feature as it leads to obtain a lower $HC_5$.

**Inclusion of uncertainty on Critical Effect Concentration**

A last group of Bayesian approaches aims at taking into account the uncertainty on the toxicity data. Noting that several values can be available for one species-contaminant pair and that the general practice is to take the geometric mean of these values,

[8] The DIC is a generalisation of the AIC used for model selection in the Bayesian framework. It includes the deviance and a penalisation for model complexity.

[Hickey and Craig, 2012] argued that there exists a large inter-test variation and proposed a model to take it into account. As it is not a simple problem, they proposed as a first step to assume that the inter-test variability was homogeneous across all contaminants and to estimate it in a hierarchical model. They found that in general, including intertest variability revealed that the classical SSD is overprotective, but there are exceptions to that rule which suggest that classical SSD might sometimes under-protect. In any case, accounting for inter-test variability is a better alternative than taking the geometric mean.

Another way to include uncertainty on the toxicity data was proposed by [Moore et al., 2010], based on the inclusion of the raw data in the SSD approach instead of the CECs, which are a only summary of the raw data (see also subsection 1.1.2). They proposed to use a hierarchical model to account for inter-replicate variation and inter-test variation of species sensitivity, by including the concentration-effect model for the raw data, but they considered only one contaminant. They assumed a normal error model for inter-replicate variation and a log-logistic concentration-effect model with four parameters. These four parameters were allowed to vary for each test and followed a normal distribution centred around a species specific mean, with a variance for inter-test variation identical for the four parameters. Finally, the four species-specific means also followed a normal distribution with a common variance but each their own mean. These means were given peaked centred normal priors ($\mathcal{N}(0, 10^{-4})$) while the variance parameters (or rather the precision parameters, i.e. the inverse of the variance) at the replicate, test and species level were given flat $gamma(0.001, 0.001)$ priors. With the fitted hierarchical they were able to generate $EC_x$ and construct an SSD including sources of variance at the level of model fit, test and species. They did not justify much their choice of prior distributions, nor the choice that the variability of the four parameter should be the same (common variance parameter). The concentration-effect model being nonlinear, the parameters might play a very different role and vary on totally different scales, justifying different variance parameters. However, their methodology provide an elegant way to deal with uncertainty on the toxicity data. This work was the starting point of much of the work developed in this thesis. The structure of the hierarchical models, the concentration-effect models, the prior distributions and the analysis of the posterior distributions used in this thesis ended up to be quite different, but the motivation for hierarchical modelling owes much the material presented in [Moore et al., 2010].

## 1.3 Motivation and outline of the thesis

**Motivation**

We drew a link between the multiple developments of SSD which used the Bayesian framework by mentioning that they all endeavoured to include external information in classical SSD, information gathered from the taxonomy of the species, from expert knowledge of from bioassays performed on other chemicals. The last development proposed by [Moore et al., 2010] fits that description in that it uses information from the raw data instead of the CECs only, which can be considered as information external to classical datasets. However, that information is not external in the same sense as taxonomic in-

formation or expert knowledge. Constructing an SSD from the raw data really amounts to avoid artificially summarizing the data, and in principle this does not require more experimental effort than to measure the CECs in the first place. Rather, the effort is switched to the modelling side. Toxicity data are generally expensive to collect and rare compared to the needs. Acute toxicity experiments are easiest to perform and already take days, while chronic toxicity experiments might last months and are possible only for very specific species whom the experimenter knows how to maintain alive for so long (this requires knowing what temperature, flow, food, physical and chemical conditions are appropriate for that species). Moreover, there is a growing consensus towards reducing animal testing which does not push in the direction of increasing data collection. Therefore, it is crucial to make the best of the available data. A lot of data is currently discarded in the classical SSD approach: 1) concentration-effect curves for which it is not possible to estimate a CEC because it is too imprecise or because no effect was observed for the concentration range tested are altogether discarded, 2) raw data are summarised at the loss of valuable information, and 3) when data are recorded over time, only the last measurement is usually taken into account.

**Outline of the thesis**

This thesis aims at revisiting SSD by proposing methods to take advantage of the data currently discarded. We show how to account for these three types of data in SSD and consider possible uses of the additional information made available. The remainder of the thesis is divided in three chapters: Chapter 2 introduces censored data and describes methods to include them in SSD. It then presents a web-tool specifically designed to include censored data in SSD and ends with a study of the added value of taking censored data into account on various examples. Chapter 3 presents a hierarchical model of SSD which includes together the concentration-effect model at the species level and the distribution structure at the community level. It explains that using this approach, the raw data is not summarised as in classical SSD and uncertainty on the CEC is adequately propagated to the $HC_5$. The chapter begins by a section investigating several possible concentration-effect models. It then presents the hierarchical structure of the model and shows how additional information extracted from the data can be used to construct a different description of the community response to contaminant exposure, which complements the $HC_5$. The results of the hierarchical SSD are compared to those of classical SSD to assess the effect of discarding uncertainty on the CEC. Chapter 4 presents a method to construct a time-resolved SSD from survival data using a hierarchical model incorporating a Toxico-Kinetic model. It begins by a description of the Toxico-Kinetic model including the derivation of the survival probability equations, followed by a description of the hierarchical structure of the time-resolved SSD. The time-resolved SSD is then compared to the classical SSD to understand how time evolution comes into play. The approach is illustrated on a salinity tolerance dataset for a large number of species from Australia and France and concludes with possible developments of the approach to compare the effects of salinity reduction intervention strategies.

Chapter 2

# Inclusion of censored data in Species Sensitivity Distribution

This thesis begins with a description of a simple method to include censored data in SSD. Censored data is usually discarded or transformed, which has a deleterious effect on SSD. We first show that it is simple to adapt the standard SSD method to include censored data, then we present a web-tool which encapsulates the method and finally, we show the added value of including censored data in SSD. The web-tool is based on an R-package which is not dedicated to SSD and the work of this thesis consisted in adapting it to SSD. It included participating in the development of this web-tool, in the decisions concerning the methodological choices and studied the influence of including censored data into SSD. This work was published in [Kon Kam King et al., 2014].

## 2.1 Different frequentist methods to deal with censored data

### 2.1.1 Definition

Censored data is a general name given to data which are not in the form of fixed values but belong to an interval, bounded or not. Censored sensitivity data occur when it is not possible to determine a single value for the CEC for a given species. Possible reasons are that 1) the highest concentration tested does not have any noticeable effect, 2) only a tiny amount of contaminant already stamps out all the individuals, 3) the measurement is simply too imprecise to be reasonably described by a single value instead of an interval. In such cases, it is only possible to give a *lower* bound, a *higher* bound or an *interval* to the CEC. Such data are said to be *right-censored*, *left-censored* or *interval-censored*,

respectively (see also Figure 2.1). Censored data can also occur when there are multiple values for the sensitivity of one species to a given toxicant. When the quality of the data seem equivalent, ECHA's advice[ECHA, 2008] is to use the geometric mean as a replacement for the different values. It might be more cautious to use these multiple values to define an interval containing the sensitivity of that species.

Censored data are very different from doubtful or questionable data obtained from failed experiments. They are produced using a valid experimental procedure and they contain information as valid as non-censored data. Censorship is very common, especially for rare species where there are scant data available and for which no standard test procedure exists. There is a downside in discarding censored data, as they could represent the better part of an extended dataset. For instance, in the work by Dowse et al.[Dowse et al., 2013], discarding censored data entails a division of the number of tested species by a factor 8. In spite of their ubiquity, censored data appear to be very much ignored in ecotoxicology. To our knowledge, there is no example of SSD including all types of censored data in a frequentist framework. It is possible in a Bayesian framework[Kefford et al., 2012a, Hayashi and Kashiwagi, 2010, Hickey and Kefford, 2008], but fitting a Bayesian model requires a certain statistical expertise. Censored data are typically either discarded or substituted with arbitrary values, which is a bias-prone approach in general[Helsel, 2006]. This can be understood because discarding left or right censored data entails excluding from the dataset only the most sensitive or tolerant species, which produces an obvious bias.

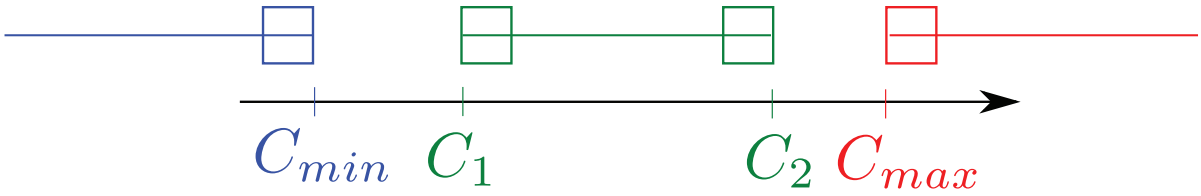



Figure 2.1: Illustration of the different types of censored data. Left censored data is in blue, interval censored data in green and right censored data in red.

### 2.1.2 Kaplan-Meier estimator for the Species Sensitivity Distribution with only left or right-censored data

One frequentist method which has been used to include censored data is based on the Kaplan-Meier estimator[Dowse et al., 2013, Kefford et al., 2006, Kefford et al., 2003, Hickey and Kefford, 2008, Kefford et al., 2005b, Kefford et al., 2012a]. It is a non-parametric estimator which can accommodate for right or left-censored data, but not for interval censored data nor for a mix a right and left-censored data. The Kaplan-Meier estimator is the non parametric maximum likelihood estimator for right-censored data[Kaplan and Meier, 1958].

We give a brief description of the Kaplan-Meier estimator in the case of an SSD with right-censored CEC. Reading this description is not essential to understanding the work presented in this thesis and it could be skipped. Let us define $F$, the CDF of the SSD and $(x_1, \ldots, x_N)$ which are either the non-censored CEC or the lower bounds of the right-censored CEC sorted by ascending order.

For a given concentration $c$, the probability of finding a concentration higher than $x_i$ and the cumulative distribution are linked by:

$$1 - F(x_i) = Pr(c > x_i) \tag{2.1}$$
$$= Pr(c > x_i | c > x_{i-1}) Pr(c > x_{i-1}) \tag{2.2}$$
$$= Pr(c > x_i | c > x_{i-1}) Pr(c > x_{i-1} | c > x_{i-2}) \dots Pr(c > x_1 | c > 0) Pr(c > 0) \tag{2.3}$$

$Pr(c > 0) = 1$ since the concentration is necessarily positive, and the probability to have a concentration higher than $x_i$ given that it is higher than $x_{i-1}$ can be estimated by:

$$\forall i, P(c > x_i | c > x_{i-1}) = \frac{n_i - d_i}{n_i} \tag{2.4}$$

where $n_i$ is the number of species whose CEC is higher than $x_i$ ($Card(\{x_j, j \in \mathbb{N}^n; x_j \geq x_i\})$) and $d_i$ is the number of CEC which are not censored and have exactly the value $x_i$. If the CEC are estimated from the fit of a model (a continuous variable), there should not be two CEC with the exact same value and most often $d_i = 1$, but $d_i$ can be greater than 1 to take into account the possibility of ties.
The Kaplan-Meier estimator for the CDF of the SSD is obtained by recurrence from Equation 2.4:

$$\hat{F}(x) = 1 - \prod_{j; x_i < x} \frac{n_i - d_i}{n_i} \tag{2.5}$$

If $x_i$ is the lower bound of a right-censored CEC and there is no tie, then $d_i = 0$, Equation 2.4 is equal to one and does not change the product of Equation 2.5. Therefore, censored CEC participate in the values $n_i$ but not in the $d_i$. This is the mechanism by which the Kaplan-Meier estimator accounts for censored data.
Using a similar construction, the estimator can be adapted for the case where there are left-censored data only[Klein and Goel, 2013], but not for interval censored data or for a mix of left and right censored data.

### 2.1.3 Maximum likelihood for all types of censored data

However, there is a simple method to include all types of censored data in a frequentist framework. Parameter estimation of a distribution on any type of censored data can be performed using a natural extension of the maximum likelihood method[Helsel, 2005]. This approach is traditionally favoured, for it benefits from a strong body of analytical works and asymptotic properties: the maximum likelihood estimate converges to the true value of the parameters (consistency), it is the fastest estimate to converge (efficiency), the difference with the true value is distributed normally (normality), which provides confidence intervals on the parameters. Moreover, other techniques such as moment matching might not be easy to use, because moments are not trivially estimated for censored data[Kroll and Stedinger, 1996]. Let $x_i$ be $N$ sensitivity data following distribution $f$ of parameter $\theta$. The likelihood function for non-censored data writes as follows:

$$L(\theta) = \prod_{i=1}^{N} f(x_i|\theta) \tag{2.6}$$

This likelihood function can be extended to censored data. Let $x_i$ be the $N_{nc}$ non-censored data, $x_j^{up}$ the $N_{lc}$ upper bounds for left-censored data, $x_k^{low}$ the $N_{rc}$ lower bounds for right-censored data and $\left(x_l^{low}, x_l^{up}\right)$ the $N_{ic}$ pairs of bounds for interval-censored data. Then, the previous likelihood function now extends to:

$$L(\theta) = \prod_{i=1}^{N_{nc}} f(x_i|\theta) \times \prod_{j=1}^{N_{lc}} \left(F(x_j^{up}|\theta)\right) \times \prod_{k=1}^{N_{rc}} \left(1 - F(x_k^{low}|\theta)\right) \times \prod_{l=1}^{N_{ic}} \left(F(x_l^{up}|\theta) - F(x_l^{low}|\theta)\right) \tag{2.7}$$

where $F$ is the CDF of distribution $f$.
We see that the likelihood function for censored data (Equation 2.7) writes as a product of four terms, the first being the likelihood for non-censored data (Equation 2.6) and the next three corresponding to the left-censored data, right-censored data and interval-censored data respectively.

## 2.2 MOSAIC_SSD

### 2.2.1 Introduction

It is possible to use the method described in the previous section using the R-package *fitdistrplus*[Delignette-Muller et al., 2013]. R-packages *survival*[Therneau, 2014] and *NADA*[Lee, 2013] offer the same possibility. However, they require a certain fluency in the R programming language, preventing the widespread use of censored data in ecotoxicology. Minitab[min, 2000] is a commercial software with a graphical user interface which fits multiple distributions to censored data rather easily, but there does not seem to be any open-source alternative. Moreover, *fitdistrplus* and these other packages and software are not specifically designed for SSD and their versatility in the choice of distributions and fitting methods might discourage inexperienced users. Thus, we developed a web-interface, MOSAIC_SSD (`http://pbil.univ-lyon1.fr/software/mosaic/ssd/`), which is a wrap up of *fitdistrplus* into a SSD-dedicated online tool. MOSAIC_SSD enables anyone to perform a simple, yet statistically sound SSD analysis including censored data without worrying about the conceptually difficult underlying statistical questions. The web-interface is easily accessible via any browser and simple to use: given an input dataset, it sends the calculation to a server then hands in the result.

### 2.2.2 Methodological choices

**Distributions**

We chose to offer few options to keep the tool more user-friendly. The user can choose one or two among the log-normal and log-logistic distributions. These two distribu-

tions are the most widely used[Wheeler et al., 2002] and parameter estimation appears robust enough to accommodate for most datasets, as they contain only two parameters. In order to select which distribution describes the data best, the first step is to perform a qualitative assessment by looking at the representative curves. The value of the likelihood function for each model can then be used as a further decision criterion. The log-logistic distribution has heavier tails than the log-normal and is therefore more conservative in the determination of the lower bound of the confidence interval on the $HC_5$[Aldenberg and Slob, 1993].

**Bootstrap confidence intervals**

The concept of bootstrap was introduced in section 1.2.2. Confidence intervals on the $HC_5$ maybe be computed via bootstrap by generating samples with the same size as the original dataset and fitting an SSD on each of them. The reasoning is that each generated sample is considered as a repetition of the experiment and that the distribution of the $HC_5$ for each sample approaches the frequentist distribution of the $HC_5$ resulting from repeated sampling. The quantiles of these $HC_5$ thus provide a confidence interval. Calculating the confidence intervals using a bootstrap method [Efron and Tibshirani, 1993] has the advantage of using a unified framework for every distribution. The samples might be generated using two methods, parametric bootstrap and non-parametric bootstrap. In the case of parametric bootstrap, the parameters of the SSD are estimated by maximum likelihood and samples are generated using these parameters and the form of the distribution. In the case of non-parametric bootstrap, samples are generated by sampling with replacement from the original data. Parametric bootstrap was chosen for non-censored datasets, because a parametric SSD it fitted on the data and because it is easier to observe convergence of the bootstrap confidence intervals with parametric bootstrap on small datasets. Non-parametric bootstrap was chosen for censored datasets, as there is no trivial method to generate a sample containing censored data[1].
Within MOSAIC_SSD, the bootstrap 95% confidence intervals are automatically computed. They yield confidence intervals on the parameters of the distribution and on several computed $HC_p$. The number of generated bootstrap samples to use for computing the confidence interval must be determined and this number is strongly dependent on the dataset. Figure 2.2 illustrates the evolution of the bootstrap confidence intervals on $HC_5$s as the number of samples is increased. For a fixed number of samples, convergence of the bootstrap procedure means that repeating the bootstrap procedure several times with the same number of samples returns the same result for the confidence interval. As the bootstrap procedure is not guaranteed to converge, I implemented an automatic check of bootstrap convergence based on the following approach: for each dataset, the bootstrap procedure is run five times with 5000 samples, comparing the magnitude of the fluctuations of the bounds of the confidence interval to the span of the confidence interval. When this fluctuation over the five runs is negligible with respect to the span of the confidence interval, the bootstrap procedure is considered converged[2] and the con-

[1] Starting from the estimated parameters, one would have to choose for each CEC generated from the distribution the type of censoring and the value of the censoring bounds. We have not found a procedure which is not strongly arbitrary and which would generate randomly censored sample.

[2] Let $(L_i)_{i\in\{1,\dots,N\}}$ be the $N$ lower bounds of the bootstrap confidence intervals and $(R_i)_{i\in\{1,\dots,N\}}$ the

fidence interval is computed with the 25000 concatenated samples. In the case where the bootstrap procedure fails to converge, additional computations are launched. If the bootstrap eventually converges, or if the process has reached the time limit, the user is advised whether the confidence intervals are reliable.

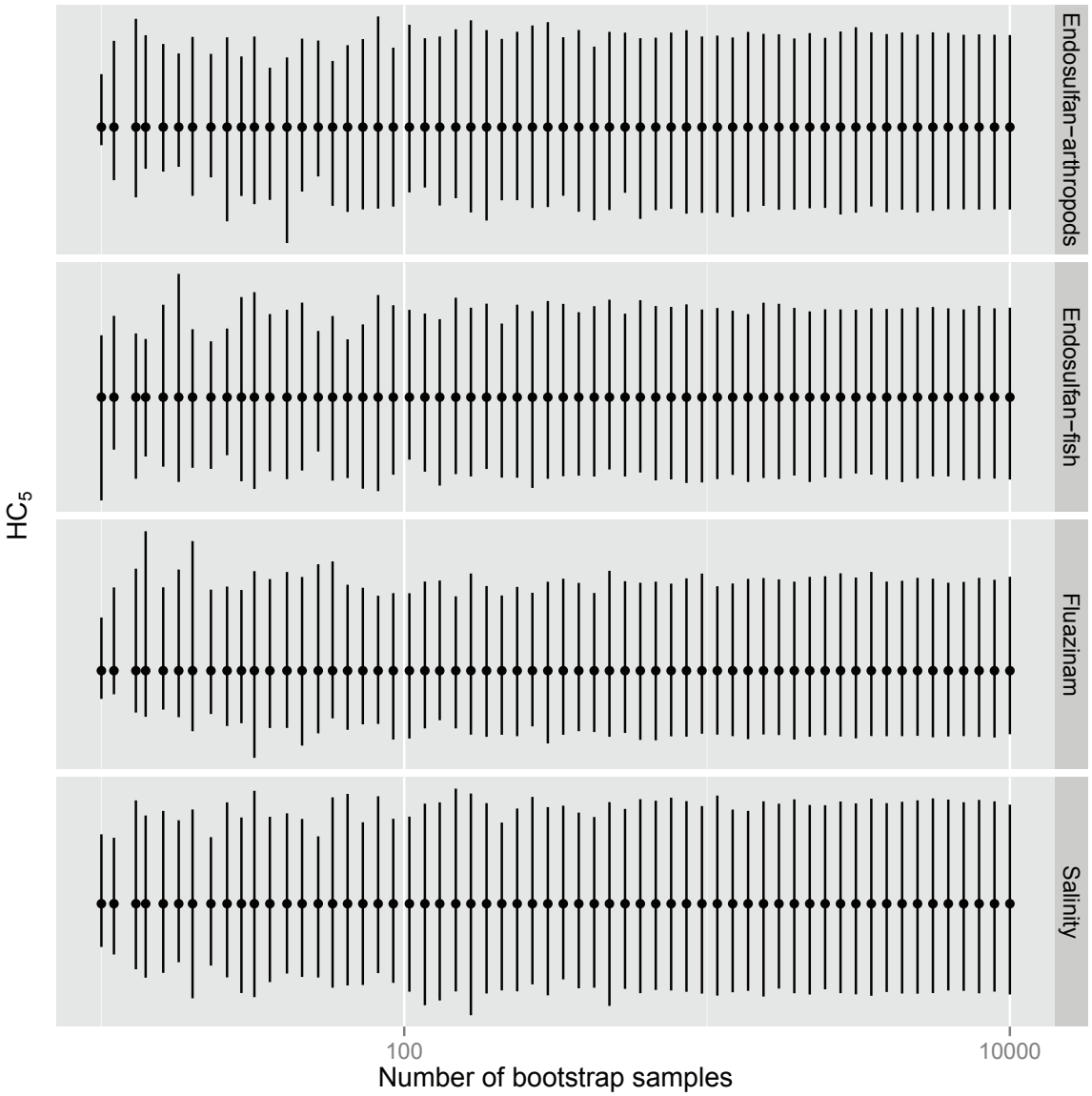



Figure 2.2: $HC_5$ estimate and bootstrap confidence interval for the four example datasets provided in MOSAIC_SSD. The units of the $HC_5$ are specific to each dataset, and the $y$ axis is in log scale.

higher bounds. Let $s_L = \sqrt{Var_i(L_i)}$ and $s_R = \sqrt{Var_i(R_i)}$. We arbitrarily consider the bootstrap converged if $\frac{1}{N}\frac{max(s_L,s_R)}{<R_i>_i - <L_i>_i} < 5 * 10^{-2}$, where the division by $N$ is intended so that the condition does not become more and more stringent with the number of bootstrap samples.

### 2.2.3 Interface

Figure 2.3 shows a screenshot of the result page of the analysis with an example dataset (provided in MOSAIC_SSD) containing censored data and documenting the salinity tolerance of riverine macro-invertebrates[Kefford et al., 2006] (hereinafter referred to as the *censored salinity* dataset). The dataset contains 72-hrs $LC_{50}$ values for 110 macro-invertebrate species from Australia. Data were collected using Rapid Toxicity Testing (RTT)[Kefford et al., 2005a] and contain non-censored, right-censored and interval-censored data. The result page shows a graphical representation of the censored data, the Hazardous Concentration for $p$% of the species ($HC_p$) computed for various values of $p$ which might be of interest and the bootstrap confidence intervals within brackets. Figure 2.3 also shows the output of an SSD analysis with a non-censored dataset. It is actually a non-censored version of the salinity dataset described earlier. The transformation from censored to non-censored dataset followed the customary approach to censored data, which consists in discarding some type of data and transforming others (more details in subsection 2.3.1). An analysis with non-censored data follows identical steps as with censored data and yields results with the same outline, except that a traditional CDF is used to represent the data.
The most apparent difference between the outputs of the censored dataset and the non-censored dataset is the representation of the CDF. For non-censored data, the CDF is represented using the traditional Hazen plotting positions[Posthuma et al., 2010]. The choice of plotting positions remains arbitrary and there is no perfect solution[Hickey and Craig, 2012, Posthuma et al., 2010], so we gave preference to the most standard approach. Representation of censored data CDF is a nettlesome problem by itself. Building a CDF implies defining an ordering for the data. If obvious for non-censored data, such an ordering has little meaning for interval-censored data. Should they be ranked according to the median of the interval? To the higher bound, the lower? What to think of left/right-censored data? Within MOSAIC_SSD, we chose to answer this problem using the Turnbull estimate of the CDF, which is a non-parametric maximum likelihood estimator of the CDF[Turnbull, 1976]. This estimate is obtained through an expectation-maximisation algorithm and yields the CDF which predicts the data with the highest probability. The Turnbull estimate is represented as a stepwise curve as on Figure 2.3 (top panel).
Finally, we intended MOSAIC_SSD as a stepping stone to perform further analysis with *fitdistrplus*. The last item on the MOSAIC_SSD result page (not shown on the screenshots) is an R script which allows to perform a similar analysis using *fitdistrplus* through a copy and paste operation in R. This script can be tuned by slightly changing some of the options : for instance, an $HC_p$ for different values of $p$ might be computed, one may try alternative distributions or a different fitting method. Moreover, this script ensures transparency and traceability of the results obtained with MOSAIC_SSD.

## 2.3 The added value of censored data

Changing a few parameters in the R script provided within MOSAIC_SSD, it is possible to use *fitdistrplus* to investigate on several fundamental aspects of SSD. We used

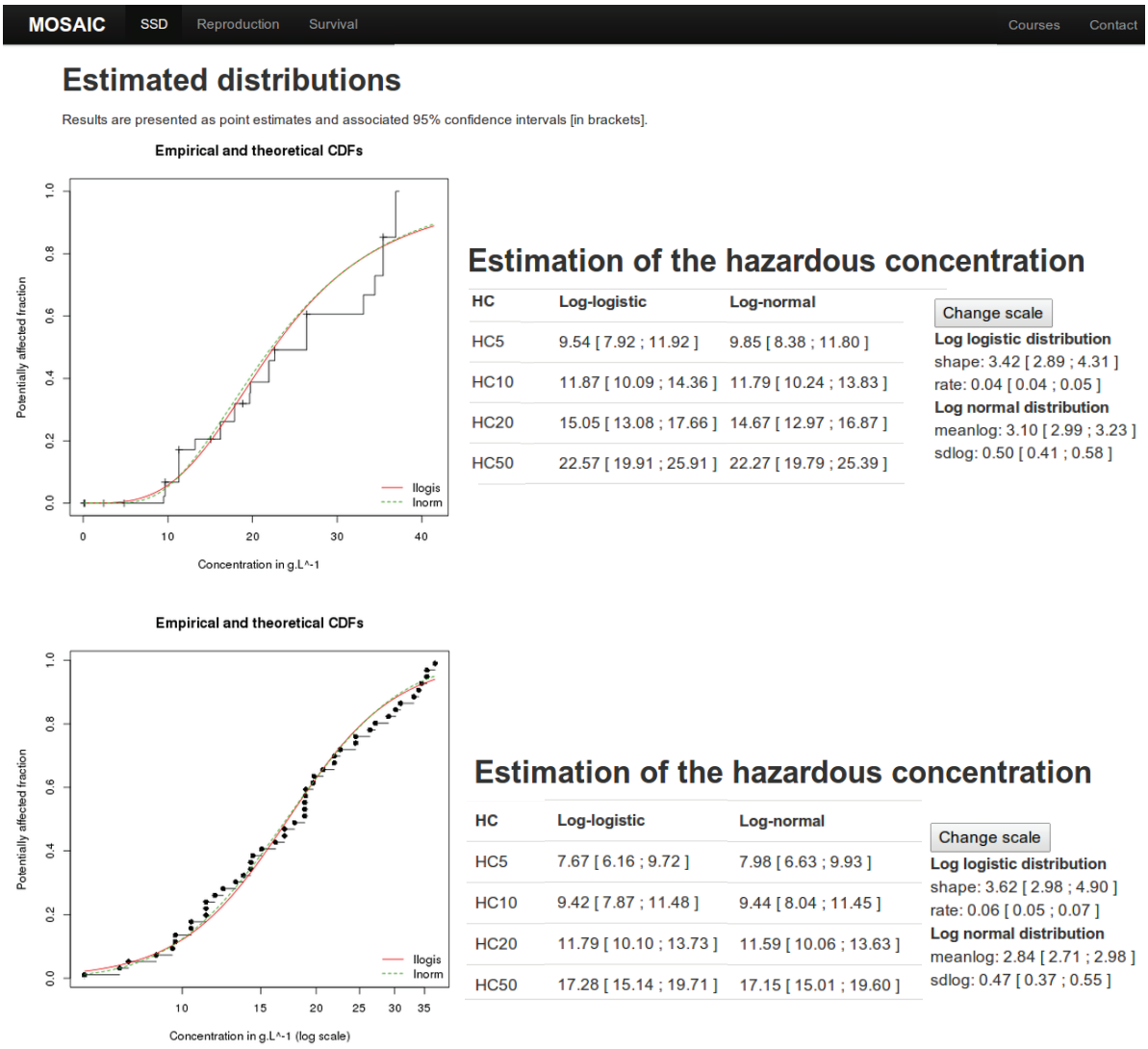



Figure 2.3: Screenshot of the result page of MOSAIC_SSD on the salinity censored dataset. The top panel shows the output of MOSAIC_SSD on the salinity dataset, the bottom panel shows the output on a non-censored dataset obtained from a transformation of the salinity dataset.

*fitdistrplus* to study the influence of including censored data on the predicted $HC_5$, and for the purpose of comparing censored to non-censored data, we changed the bootstrap method for non-censored data from parametric to non-parametric to remove a potential confounding factor. A customary approach when dealing with censored data is to discard or to transform it. More precisely, it is frequent to discard left or right-censored data and to take the middle of the interval-censored data as a single value. Two datasets were analysed to assess the effect of such data transformation on the predicted hazardous concentrations.

### 2.3.1 Two examples

In the censored salinity dataset mentioned earlier, out of 108 $LC_{50}$, 89 (82.4%) are censored, among which 60 (55.6%) are right-censored and 29 (26.8%) interval-censored. Most of the censored data resulted from testing of rare species, for which the small number of individuals captured prevented the calculation of an $LC_{50}$ by fitting a concentration-response model[Kefford et al., 2006]. This extensive dataset was collected to be as representative as possible of the species found in nature[Kefford et al., 2006]. Hence, a first asset of taking censored data into account is to abstain from discarding or altering the vast majority of the data. The resulting SSD is therefore more representative of the community it aims to describe. Moreover, using only non-censored data in the analysis introduces a strong selection bias towards abundant species. This is particularly problematic when some rare species are likely to be among those that the environmental manager wishes to protect by carrying an SSD analysis. The second dataset was published by Koyama et al. [Koyama, 1996], and contains vertebral deformity susceptibilities of marine fishes exposed to trifluralin (hereinafter referred to as the *censored trifluralin* dataset). The measured endpoint are 96-hrs $EC_{50}$ on 10 species. Four of the $EC_{50}$ are censored, among which two are right-censored and two are left-censored. On this dataset, the obvious advantage of taking censored data into account in that it allows to fit an SSD on 10 species, whereas discarding the censored data reduces the size of the dataset to six species only, which is below the minimum recommendation of ECHA (of 10, preferably 15[Aldenberg and Rorije, 2013]).
A non-censored version of the two datasets (hereinafter referred to as the *transformed salinity* and *transformed trifluralin* datasets) was obtained following the habitual procedure of discarding the right or left-censored data and taking the middle of the interval-censored data. Fitting the log-normal distribution on the censored and transformed versions of the datasets showed that discarding censored data had an adverse effect on the predicted $HC_5$ (Figure 2.4). For the salinity dataset, discarding the right-censored data induced a clear upward bias for the cumulative curve and a therefore greater $HC_5$ (Figure 2.4 left). The estimates for the $HC_5$ were: 9.85 $g.L^{-1}[8.38; 11.80]$ for the censored dataset and 7.98 $g.L^{-1}[6.63; 9.93]$ for the transformed dataset, respectively. An unnecessary high hazardous concentration might seem a harmless error, for using the transformed salinity dataset would have proven more protective. However, that incorrectly low value might motivate the use of costly decontamination measures at a specific location, when efforts could be spared and distributed elsewhere.
The influence of censored data is dataset-dependent and the bias could be in the opposite

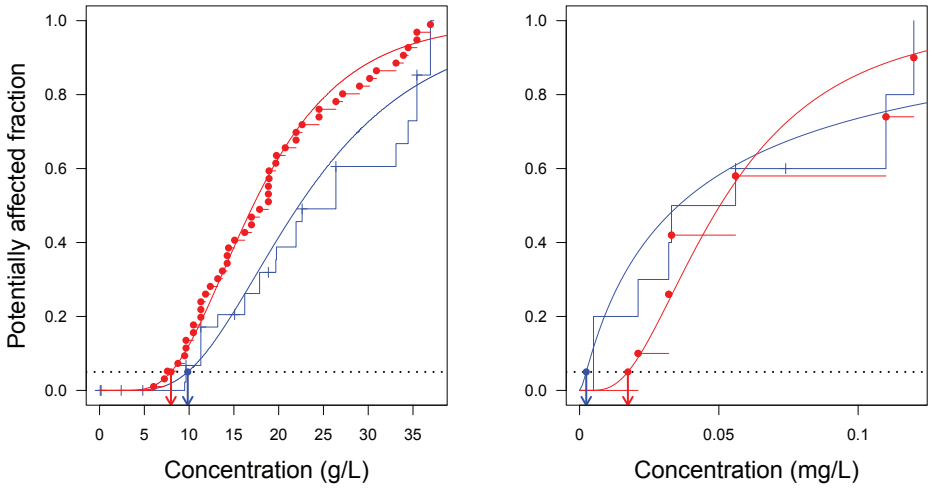



Figure 2.4: Fitted and empirical cumulative distribution and the $HC_5$ for the salinity dataset (left) and the trifluralin dataset (right). The dotted line corresponds to a potentially affected fraction of 5%. Vertical arrows indicate the $HC_5$. The blue line is for the censored dataset, the red for the transformed dataset. A log-normal distribution was fitted on the dataset.

direction. This is illustrated on the trifluralin dataset (Figure 2.4 right). Fitting the log-normal distribution yielded the following estimates for the $HC_5$: $2.4 \times 10^{-3}$mg.L$^{-1}$[$4.7 \times 10^{-5}$; $2.6 \times 10^{-2}$] for the censored dataset and $1.7 \times 10^{-2}$mg.L$^{-1}$[$8.9 \times 10^{-3}$; $4.3 \times 10^{-2}$] for the transformed dataset, respectively. Discarding the censored data led to underestimate the variability in the community sampled by the tested species. Therefore, the width of the distribution was underestimated and the fifth percentile had a value too large. Discarding the censored data led to an underestimation of the trifluralin real toxicity and of its potential hazard to the environment. Another striking differentiation was that the span of the confidence interval was much larger when censored data were included in the SSD. It reveals that a possible effect of transforming censored datasets is to severely underestimate the width of the confidence interval and to give overconfident predictions on the hazardous concentrations.

### 2.3.2 RIVM database

In order to confirm these findings, we studied a published dataset containing 3442 contaminants and 1549 species[Hickey et al., 2012]. This dataset contained both censored and non-censored data, thus we compared the $HC_5$ obtained on the non-censored dataset to the $HC_5$ obtained on the transformed dataset including only censored data. In order to have consistent endpoints for SSDs, we restricted the dataset to $EC_{50}$ only. To ensure that there was at least a little variation between the complete and transformed dataset, we also chose to focus on chemicals for which the proportion of censored data was superior to 10%.

**Data aggregation**

As the dataset may contain several sensitivity values for a given (contaminant, species, endpoint) set, the data was aggregated according to the following scheme: to define the censored sensitivity value for a species, the largest interval containing all the data (censored or not) was chosen, i.e. the smallest lower bound and largest higher bound. This is a precautionary approach, as it defines an interval which ought to contain the true sensitivity value. To define the non-censored toxicity data for a species and endpoint, all the interval censored data were transformed to non-censored data by taking the middle of the interval. Left or right censored data were discarded. The unique non-censored toxicity value was then chosen as the geometric mean of all the non-censored sensitivity values for that species and endpoint, as recommended by ECHA[ECHA, 2008].

**Description of the subsets**

Two subsets of the dataset were considered: 1) a well documented subset (data-rich), containing $EC_{50}$ and more than 10 species per contaminant in order to produce good quality SSDs. This first subset comprised 239 contaminants. 2) A poorly documented contaminant dataset (data-poor), containing $EC_{50}$ and between 5 and 9 non-censored species sensitivity values after aggregation. This second subset comprised 180 contaminants. This second subset represents a common situation when there is insufficient precise information on the species sensitivities to the contaminant.
On each of these subsets, we studied the ratio of the *non-censored* $HC_5$ over the *censored* $HC_5$. We also studied the ratio of the *non-censored* lower bound of the 95% confidence interval on the $HC_5$ ( $HC_{5,2.5\%}$ ) over the *censored* $HC_{5,2.5\%}$ for both datasets. The $HC_{5,2.5\%}$ has been proposed to derive safe concentration levels for contaminants[Wheeler et al., 2002].

**Calculation of the confidence intervals**

The confidence intervals were computed using 50001 bootstrap samples. We used non-parametric bootstrap (resampling from the data) so we could use the same method for censored and non-censored data. The appropriate number of bootstrap samples was determined by repeating the analysis several times and checking that the quartiles and extremal values of the ratios $\frac{HC^{ncens}_{5,2.5\%}}{HC^{cens}_{5,2.5\%}}$ were reasonably stable. The large amount of contaminants prevented checking individually the convergence of the confidence intervals. However, the extremal values which were used in the analysis were guaranteed to have converged, since they were directly observed.

**Results of the analysis**

The study of the large ecotoxicity database revealed that it is risky to discard or transform censored data. In both subsets, the non-censored data led to an overestimation of the $HC_5$ and of the $HC_{5,2.5\%}$ in roughly half of the cases, showing that the bias induced by the discarding and transformation of the data can be in both directions. Figure 2.5 shows that for a large proportion of the contaminants, the change in $HC_5$ and $HC_{5,2.5\%}$ was rather small. Part of the explanation is that the proportion of censored data is modest for a good proportion of the contaminants (the $8^{th}$ decile is at 50% of censored data). But this also demonstrates that in many cases, discarding and transforming the censored data will not have a very strong impact on the $HC_5$ and the $HC_{5,2.5\%}$. However,

since the $HC_5$ or the $HC_{5,2.5\%}$ are used for risk assessment, it is legitimate to consider the worst case scenario, i.e. the worst bias on the $HC_5$ and the $HC_{5,2.5\%}$. This represents an estimation of the bias one should be ready to accept when transforming censored data into non-censored data. On the data-rich subset, non-censored data lead to predict $HC_5$ and $HC_{5,2.5\%}$ between 5 times too large and 4 times too small (Table 2.1). Such bias factors could be compared to the safety factors applied to define safe concentration levels, which range from 1 to 5[ECHA, 2008]. The maximum bias is comparable to these safety factors and it could be argued that they exist indeed to compensate for this sort of bias. However, the bias can be much worse on the data-poor subset, where non-censored data lead to predict an $HC_5$ up to 80 times too large, and an $HC_{5,2.5\%}$ overestimated by several orders of magnitude(Table 2.1). Three contaminants had a bias factor on the $HC_5$ greater than 5. Five contaminants had a bias factor on the $HC_{5,2.5\%}$ greater than 5. The result of this analysis is that when sufficient non-censored data is available, the value of the $HC_5$ might prove relatively insensitive to the degradation of the information induced by arbitrarily transforming censored data into non-censored data. When little data is available however, it appears crucially important to include censored data in the determination of the $HC_5$. The bias on the $HC_5$ can lead to an overestimation by a factor 80, yielding safe concentrations which fail to protect the target communities. The bias on the $HC_{5,2.5\%}$ can even be greater, as it reached five orders of magnitude on the dataset studied.

Table 2.1: Extremal $\log_{10} HC_5^{ncens}/HC_5^{cens}$ and $\log_{10} HC_{5,2.5\%}^{ncens}/HC_{5,2.5\%}^{cens}$ ratios for the two subsets, with the corresponding bias factor in parentheses. The bias factor corresponds to the maximum overestimation ($\times$) underestimation ($\div$).

| Subset | Data-rich | Data-poor |
|---|---|---|
| Number of contaminants | 239 | 180 |
| $\max(\log_{10} HC_5^{ncens}/HC_5^{cens})$ | $0.37(\times 2.34)$ | $1.92(\times 82.22)$ |
| $\min(\log_{10} HC_5^{ncens}/HC_5^{cens})$ | $-0.61(\div 4.04)$ | $-0.61(\div 4.04)$ |
| $\max(\log_{10} HC_{5,2.5\%}^{ncens}/HC_{5,2.5\%}^{cens})$ | $0.7(\times 4.98)$ | $5.45(\times 2.82 \times 10^5)$ |
| $\min(\log_{10} HC_{5,2.5\%}^{ncens}/HC_{5,2.5\%}^{cens})$ | $-0.47(\div 2.92)$ | $-0.72(\div 5.23)$ |

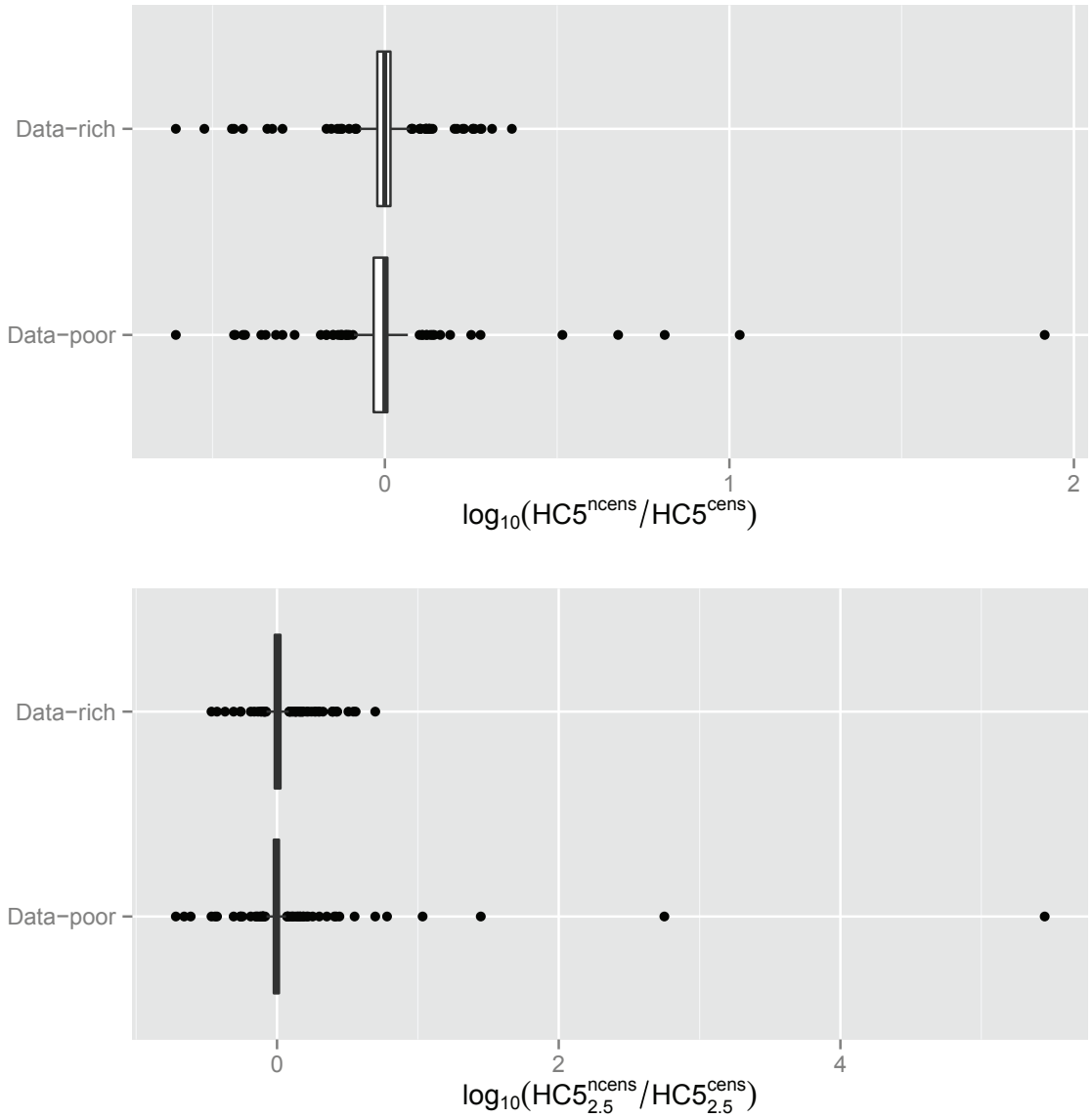



Figure 2.5: Boxplots for the log-ratio of $HC_5$ $\log_{10} HC_5^{ncens}/HC_5^{cens}$ and the log ratio of the lower bound of the $HC_5$ confidence interval $\log_{10} HC_{5,2.5\%}^{ncens}/HC_{5,2.5\%}^{cens}$. The box extends to the inner quartiles, the whiskers to 1.5× inter-quartile range as recommended in [McGill et al., 1978].

### 2.3.3 Discussion

In the introduction and the first part of the thesis, we reviewed the general approaches to fit an SSD to sensitivity data and explained how it was possible to use maximum likelihood to include censored data in SSD. We presented MOSAIC_SSD, a web-tool developed during the thesis which allows any user to perform an SSD analysis including censored data with few very simple steps. MOSAIC_SSD is an interface to a more versatile tool, the R package *fitdistrplus*[Delignette-Muller et al., 2013] and presents a

restricted number of options for an easier use. We supported the methodological approach behind MOSAIC_SSD with several arguments and showed the added value of including censored data into the SSD. Discarding or transforming censored data has been shown to alter the results of the SSD analysis. Using MOSAIC_SSD is a convenient way to take censored data into account in the fitting of an SSD, but the sound general statistical approach adopted is also an asset to perform any sort of SSD. Considering the choice of a distribution, MOSAIC_SSD provides by default two standard distributions, the log-normal and log-logistic, but it fosters the use of alternative distributions via providing a stepping stone to using the R package *fitdistrplus*. The question "which distribution best fits a dataset?" cannot have a general answer and must be addressed by testing several options. Therefore, having room to try multiple distributions is a valuable asset. Someone might wish to fit a distribution that best describes the *tails* of the dataset, because determining a $HC_5$ is an extreme quantile estimation problem. In that case, one might try a heavy tailed distribution such as the Weibull or exponential. In selecting a distribution, special care should be taken not to pick a distribution with too many parameters. One of the easily accessible software for SSD is BurrliOZ[Campbell et al., 2000], which fits the Burr III distribution using maximum likelihood and computes confidence intervals using bootstrap. The Burr III distribution is very flexible[Shao, 2000, Fox, 2008], but it contains one parameter more than the log-normal or log-logistic distributions. Fitting of a distribution with many parameters requires a lot of data and the Burr III distribution is likely to suffer from strong structural correlation among the parameters[Shao, 2000], which means that convergence of the likelihood maximisation algorithm might be difficult and the estimates produced might not be very reliable. However, BurrliOZ is currently being developed so as to fit the log-logistic distribution on small datasets, and provide a comparison between at least the log-logistic and the Burr III distribution for larger datasets[Warne et al., 2013].
We hope that putting MOSAIC_SSD at everyone's disposal will encourage the inclusion of censored data in SSD analysis as a means to make better use of all the data at hand. We did not address all the issues pertaining to the SSD approach and merely tried to improve on the existing methods, with the aim to make better use of the available data considering the cost of collecting them. There remain interrogations as to what might happen if the proportion of censored data is too great and the dataset is small. It is not possible to test this situation thoroughly, for there are many ways to censor data and no trivial way to choose between them or scan even roughly the space of possible configurations. A good practice would be to consider the span of the confidence interval around the hazardous concentration of interest and decide if the dataset is adequate for predicting such concentrations or if more data need to be collected. Taking censored data into account would therefore be crucially important to have a precise assessment of the confidence interval, and not an artificially reduced estimation as for the trifluralin dataset.
We mentioned that censored data might represent an important part of any dataset and that MOSAIC_SSD could be profitably used on many occasions. However, this work could have a more general scope, since fundamentally all data with a confidence interval could be considered as interval-censored data. Indeed, the confidence interval around an $EC_{50}$ or any CEC estimate might be considered as the range which has a 95%

probability of containing the real value and be reported as an interval-censored data. Using the confidence intervals on the CECs as censored data would provide a crude way to propagate the uncertainty on the CEC into the SSD, a fundamental problem of SSD which is seldom tackled[Dixon, 2007], but of interest nonetheless[Warne et al., 2013]. Moreover, Lowest Observed Effect Concentration (LOEC) data, which are often reported, are indeed left-censored data. The only information LOEC carries is that the NOEC lies below this concentration[Delignette-Muller et al., 2011]. Therefore, the SSD approach we propose, which includes censored data, would allow ecotoxicologists to make more of the available experimental data they used to calculate the NOEC.
However, we reach the limits of a traditional SSD based on CECs and still discard a lot of information. Indeed, a CEC is only a summary of a full concentration-effect curve. This summary sets aside several aspects of the response of a species to a contaminant, such as the slope of the curve (is the species gradually affected or is there a threshold effect ?). It is possible to include all the information present in the experimental concentration-effect curve in the SSD by building a hierarchical model of SSD. This hierarchy would model the joint probability of all the parameters describing a concentration-effect curve, not only the CEC as in the classical SSD[Moore et al., 2010]. Moreover, this would also allow to take proper account of the uncertainty on the species response modelling and to propagate uncertainty into the SSD. This is the subject of the next part of the thesis.

Chapter 3

# Hierarchical modelling of concentration-effect data

The first part of the thesis was dedicated to explaining how to take censored data into account and how failure to do so affected the quality of the SSD. In this second part of the thesis, we will focus on the fact that data from a full bioassay experiment are summarized by a single value, often given without uncertainty. This creates a number of flaws for SSD which we will try to correct by modelling the raw data instead of just a summary. We will build a global hierarchical model including the concentration–effect model together with the distribution law of the SSD. We will revisit the current SSD approach to account for more sources of variability and uncertainty into the prediction than the traditional analysis and to assess a global response for the community. Working within a Bayesian framework, we will be able to compute an SSD taking into account the uncertainty from the original raw data. We will also develop a quantitative indicator of a global response of the community to the contaminant. We will illustrate this methodology on the study of the toxicity of six herbicides to benthic diatoms from Lake Geneva, based on the biomass as endpoint. This work was developped in collaboration with the INRA[a] of Thonon les Bains and published in [Kon Kam King et al., 2015].

[a]Institut National de la Recherche Agronomique

## 3.1 Three shortcomings of Species Sensitivity Distribution

The classical SSD approach described in the introduction of the thesis and its many variants present a number of flaws[Forbes and Calow, 2002, Power and McCarty, 1997]

ranging from ecotoxicological concerns (use of laboratory data to predict field effects, inferring community sensitivity from monospecific sensitivities, chronic vs. acute effects ...) to statistical issues (fitting a distribution on a small dataset, distributional assumptions, treatment of the uncertainty, etc.). This work focuses on several of these:

**Uncertainty**

The classical SSD approach does not propagate the uncertainty on the CEC to the prediction. This is a source of concern, because following this approach, the uncertainty on the $HC_p$ depends on the number of species, but not on the quality of the data used. Several sources of uncertainty enter at the various steps of the SSD approach and all have an influence on the predicted $HC_p$ value. Firstly, there is an uncertainty on the estimate of the CEC from the experimental data: when the CEC is estimated from a concentration-effect curve or more generally from any model, it comes with a confidence interval. Secondly, uncertainty arises from the fitting of a distribution to the CECs: parameters of the distribution also have their own confidence intervals. This adds to the total uncertainty on the $HC_p$. The uncertainty of this second step has already been studied and methods have been found for specific distribution laws[Aldenberg and Slob, 1993, Aldenberg and Jaworska, 2000, Wagner and Lokke, 1991]. For other types of distributions, it is possible to use bootstrap. This uncertainty was also investigated with non parametric approaches in the estimation of the SSD[Jagoe and Newman, 1997, A.M Verdonck et al., 2001, van der Hoeven, 2001, Grist and Leung, 2002]. However, there are currently very few attempts to include together all the sources of uncertainty into the final prediction of the SSD[Aldenberg and Rorije, 2013].

**Information**

As a summary of the concentration-effect curve, the CEC retains only a fraction of the information originally present in the data. Since the aim of SSD is to model the variability in sensitivity in the community, it should be important to consider all the information available in the data to obtain the best estimation of that variability. Indeed, there is relevant biological information in all the parameters of the concentration-effect curve and their potential correlations. That information is discarded in classical SSD, although the variability on the other parameters might be as important as the variability of the CECs.

**Interpretability**

Providing an $HC_p$, the classical SSD approach outputs information about a structural response of the community only. It essentially yields the proportion of affected species for a given concentration in contaminant. It does not give information about the global response of the community[Forbes and Calow, 2002, Kefford et al., 2012b, De Laender et al., 2008], i.e. a response of the same nature as the measured endpoint. For instance, when using $EC_{50}$ for biomass reduction as input, the SSD does not say anything about the change in the biomass of the community. In other words, the SSD aims to protect the structure of the community, but does not consider the effect on the community endpoint linked to the tested species which could be growth, reproduction, biomass, respiration, photosynthesis or any ecosystem process.
The $HC_p$ represents the concentration which affects p% of the community. The term "*affect*" is directly linked to the type of CEC in terms of level of effect (for example the

$x$ of the $EC_x$) and of biological effect (lethal, non-lethal, acute, chronic). If NOEC or NEC were true no effect concentrations, one would expect the $HC_p$ to leave $(100 - p)\%$ of the community species completely unharmed. Using $EC_{50}$ however, which is a level of effect commonly selected, one expects $(100-p)\%$ of the community to remain unaffected, which means that they suffer a reduction of less than 50% to their measured endpoint. But it is not possible to determine the reduction suffered by the unaffected species, which could lie anywhere between 0 and 50 %.
The $HC_p$ for small $p$, such as the $HC_5$, is ultimately used as a risk indicator. It is compared to the actual concentration of contaminant in an environmental setting to determine if the community living there is at risk, or to define an acceptably safe concentration for that community.
To address such issues, we revisited the current SSD approach to account for more sources of variability and uncertainty into the prediction than the traditional analysis and to assess the risk for the community from a global point of view. For this purpose, we built a hierarchical model inspired by [Moore et al., 2010] including the concentration-effect model together with the distribution law of the SSD. From this hierarchical model, we were able to develop : 1) an indicator for the global response of the community, which we compared to the structural response predicted by the classical SSD; and 2) an SSD calculated on any level of effect ($x$ of the $EC_x$) including correlation among the parameters of the concentration-effect model and the uncertainty from the original data.

## 3.2 Diatom data and concentration-response model

### 3.2.1 Description of the data

This work was developed on a previously published dataset[Larras et al., 2012] containing 11 diatom species exposed to six herbicides : atrazine, terbutryne, diuron, isoproturon, metolachlor and dimethachlor. Between five and ten species were tested per herbicide (details on Table 3.1). Benthic diatoms are unicellular microalgae which form a group of high diversity (see Figure 3.1) and which are often used to monitor water quality. They are well known to evolve in the biofilm matrix, at the interface of water column and substrata. The chosen diatom species were representative of Lake Geneva benthic diatoms communities and covered a great diversity in terms of taxonomy, morphology, herbicide sensitivity and ecological traits. More details about chosen diatoms are presented in [Larras et al., 2012]. Then, a panel of herbicides was selected regarding their occurence in Lake Geneva, their hazard to microalgae and their mode of action. Atrazine and terbutryn (triazine family) and diuron and isoproturon (phenylurea family) prevent the photosynthesis at the level of the photosystem II, but with different mecanisms. Metolachlor and dimethachlor (chloroacematide family) inhibit especially the biosynthesis of very long chains of fatty acids. The sensitivity of the species was determined assessing the growth over four days as endpoint, based on chlorophyll a fluorescence (the part of light which is absorbed by chlorophyll molecule and then re-emitted at a defined wavelength), a proxy of the biomass. Bioassays were conducted in triplicates on diatom strains in their exponential growth phase, when the daily growth ratio is approximately constant. Seven to ten herbicide concentrations were tested. Chlorophyll

Table 3.1: Species tested for each herbicide.

| | Atrazine | Terbutryn | Diuron | Isoproturon | Metolachlor | Dimethachlor |
|---|---|---|---|---|---|---|
| *CRAC*[1] | × | × | × | × | × | |
| *NPAL* | × | × | × | × | | |
| *UULN* | × | × | × | × | | |
| *MAFO* | × | × | × | × | × | |
| *SEMN* | × | × | × | × | × | |
| *GPAR* | × | × | × | × | × | × |
| *FRUM* | × | × | × | × | × | |
| *CMEN* | × | × | × | × | × | × |
| *ADMI* | × | × | × | × | × | × |
| *FCVA* | × | × | × | × | × | × |
| *ESLE* | | | | | | × |

*CRAC: Craticula accomoda, NPAL: Nitzschia palea, UULN: Ulnaria ulna var. acus, MAFO: Mayamaea fossalis var fossalis , SEMN: Eolimna minima, GPAR: Gomphonema parvulum, FRUM:Fragilaria rumpens, CMEN: Cyclotella meneghiniana, ADMI: Achnanthidium minutissimum, FCVA: Fragilaria capucina var vaucheriae, ESLE: Encyonema silesiacum.*

a fluorescence was measured using Fluoroskan (Fluoroskan Ascent, Thermo-scientific, Finland) at the beginning and at the end of the experiment.

### 3.2.2 Choosing the concentration-effect model

**Representation of the original data**

The response of each set (herbicide, species, replicate) was defined as the ratio:

$$R = \frac{\beta_f}{\beta_0} \tag{3.1}$$

where $R$ is the response, $\beta_f$ the fluorescence after four days and $\beta_0$ the initial fluorescence. Taking the logarithm of R and dividing by four, it would represent the daily exponential fluorescence growth rate, a proxy for the daily biomass exponential growth rate. Given the small number of replicates and given that this was not the focus of this work, we chose not to model the replicate-effect. Therefore, data from the three replicates were assumed to come from a single experiment and inter-replicate variability was merged with measurement uncertainty. A first remark on the original data is that there seemed to be an important heteroskedasticity. Inter-replicate variability seemed larger for large values of the biomass growth rate than for small values (Figure 3.2). A second observation is that for several pairs of species, there seemed to be an observable *hormesis*[Calabrese, 2009, Kefford et al., 2008, Calabrese and Baldwin, 2003, Calabrese and Blain, 2011, Calabrese, 2005, Larras et al., 2012] phenomenon (NPAL, CRAC, Figure 3.2). For the diatom dataset, hormesis describes the potential increase in biomass growth for small concentrations which comes as a surprise given that the herbi-

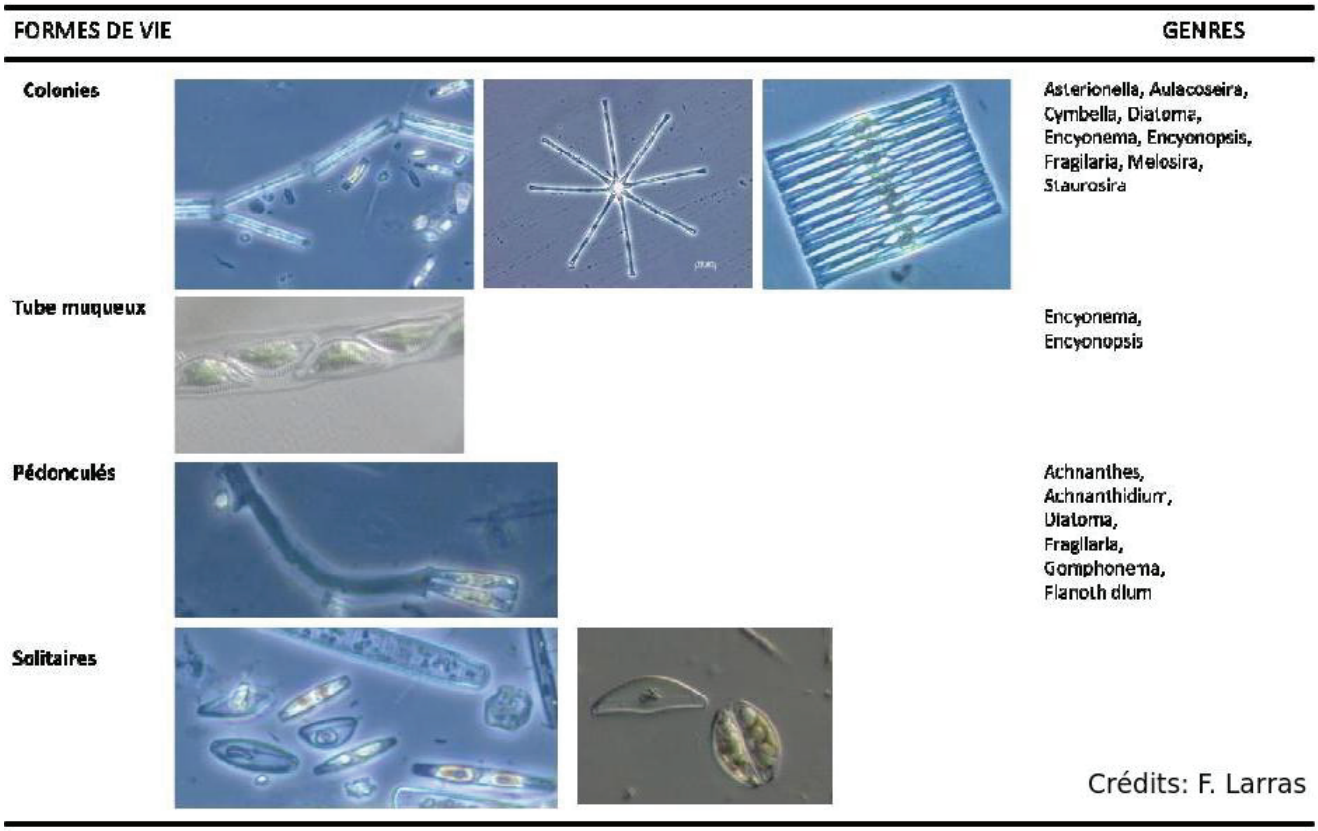



Figure 3.1: Several diatom species under the microscope.

cides are expected to inhibit growth, and indeed do so at larger concentrations. Hormesis is a very common phenomenon characterised by low-dose stimulation and high-dose inhibition[Calabrese and Baldwin, 2003], but it is apparent only when concentrations small enough are tested in the experimental setting. Hormesis challenges the definition of $EC_x$ as stimulation can either be considered as a positive or negative effect depending on the context (stimulation of algae growth might have an undesirable effect on an ecosystem). We considered the possibility to account for hormesis in the original data through the concentration-effect model.

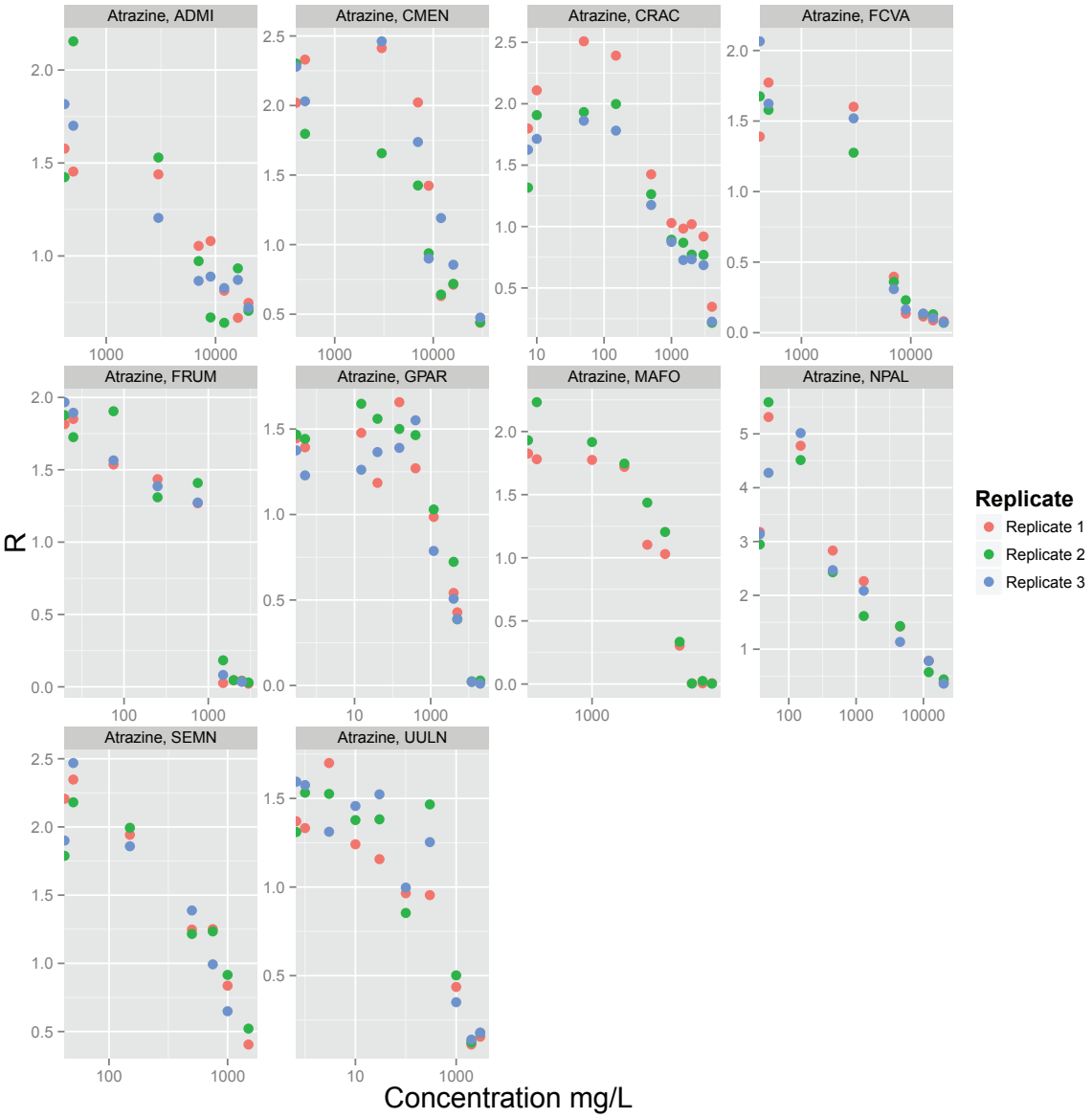



Figure 3.2: Original data for all species and for atrazine. $R$ is plotted against the concentration, with one colour per replicate. The concentration is in log-scale and the control measurements (zero concentration) are drawn on the left border of the plot. The data for the other herbicides can be found on Figure C.1.

The observed heteroskedasticity prompted a log-transformation of the data to stabilise the variance. Plotting inter-replicate variance against mean over replicates (Figure 3.3) indeed revealed a clear heteroskedasticity in the original data which was reduced to some degree by the log-transformation. Therefore, we chose to perform non-linear regression on the log-transformed data rather than on the original data.

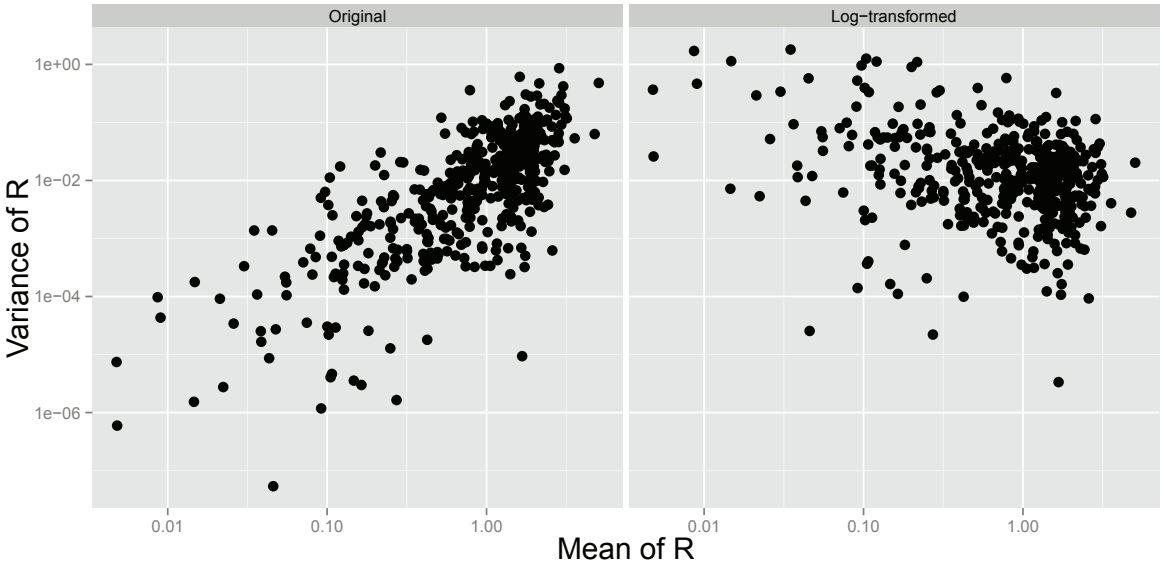



Figure 3.3: For all species and herbicides, inter-replicate variance against mean over-replicates for $R$ with the original data (left) and the log-transformed data (right). Axis follow a log-log scale.

### Possible concentration-effect models

We tested the three concentration-effect models mentioned in [Cedergreen et al., 2009] to describe the data. The first is one of the commonly used log-logistic (LL) concentration-effect model, while the other two are simple alternatives which can account for hormesis, the Brain-Cousens (BC)[Brain and Cousens, 1989] model and the Cedergreen-Ritz-Streibig (CRS)[Cedergreen et al., 2009] model. The log-logistic model is nested in the BC and in the CRS models. All three models have a parameter which describes the response at infinite concentration. We fixed this parameter to 0 as the fluorescence is expected to become null for infinitely large concentrations.
In the three-parameter log-logistic model, the response of a species $j$ to a given herbicide at concentration $C$ is:

$$R = \frac{d}{1 + \left(\frac{C}{e}\right)^b} \tag{3.2}$$

where $C$ stands for the concentration. $d$ is the response in the control. $e$ is the $EC_{50}$ in this model, ie. the concentration which induces a reduction of 50% with regards to the response in the control. $b$ is a shape parameter. Parameter $b$ is usually called the slope of the concentration-effect curve, although the real slope at $C = e$ is $-\frac{d}{4e}b$. Note that through a simple algebraic transformation, an $EC_x$ for any value of $x$ can be considered

as a parameter of the model[2]. The model is defined for $b, C, d, e > 0$ and it is decreasing with increasing concentrations with an asymptotic limit to 0. In the presence of hormesis, non-linear regression provides estimates for $d$ which can be higher than the response in the control. This is an artefactual consequence of the model not allowing for an increase with the concentration and the problem can be fixed by estimating parameter $d$ as the mean of the three control replicate, then estimating the other parameters by non-linear regression.
In the four-parameter BC model, the response of a species $j$ to a given herbicide at concentration $C$ is:

$$R = \frac{d + fC}{1 + \left(\frac{C}{e}\right)^b} \tag{3.3}$$

where $C$ stands for the concentration. $d$ is the response in the control. $e$ is **not** the $EC_{50}$ in this model ! There is no closed-form expression for the $EC_{50}$ which has to be computed numerically. $b$ is a shape parameter. $f$ is a parameter which controls the magnitude of the hormesis effect. The model is defined for $C, d, e, f > 0$ and for $b > 1$ so that it is decreasing for large concentrations with an asymptotic limit to 0.
In the five-parameter CRS model, the response of a species $j$ to a given herbicide at concentration $C$ is:

$$R = \frac{d + f\exp(-\frac{1}{C^\alpha})}{1 + \left(\frac{C}{e}\right)^b} \tag{3.4}$$

where $C$ stands for the concentration. $d$ is the response in the control. $e$ is still not the $EC_{50}$ in this model. There is again no closed-form expression for the $EC_{50}$ which has to be computed numerically. $b$ is a shape parameter. $f$ is a parameter which controls the magnitude of the hormesis effect. $\alpha$ is another shape parameter which controls the steepness of the hormetic rise as a function of the concentration. The larger the value of $\alpha$, the steeper the rise. The model is defined for $b, C, d, e, f; \alpha > 0$ and it is decreasing for large concentrations with an asymptotic limit to 0. Cedergreen et al.[Cedergreen et al., 2009] do not recommend estimating the $\alpha$ parameter from the data. Fitting the model with 6 parameters by frequentist likelihood maximisation can prove difficult. Instead, they suggest fixing it at one of the following values: $\alpha \in \{0.25, 0.5, 1\}$ to avoid over-parameterisation. These values give a different smoothness to the concentration-effect curve, and the authors suggest selecting the most appropriate[Cedergreen et al., 2009]. There were no strong reason to select one of these values for our dataset, so we settled for $\alpha = 0.25$ for the purpose of comparison as it yielded the smoothest curves (Figure 3.4).

[2] $\log \mathrm{EC}_x = \log e + \frac{1}{b}\log\left(\frac{x}{1-x}\right)$

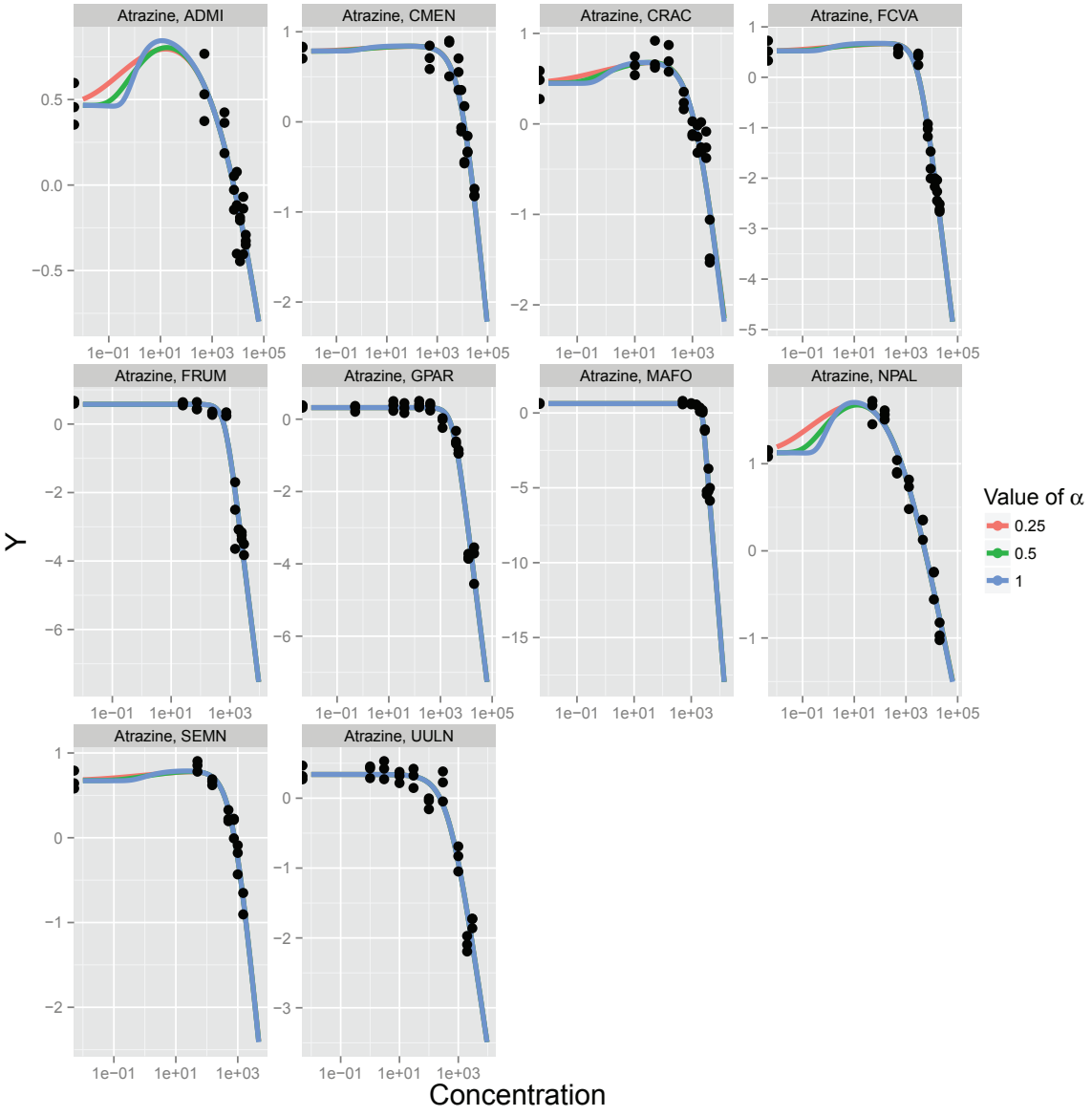



Figure 3.4: Comparison of the fit of the CRS model (Equation 3.4) on the log-transformed atrazine data for the three values of $\alpha$ suggested in [Cedergreen et al., 2009]. .

### Comparison of the different models

With the log-transformation, the following error model for the non-linear regression was used:

$$\mathrm{Y} = \ln(\mathrm{R}) = g(C, \theta_g) + \epsilon \tag{3.5}$$

where Y is the natural logarithm (ln) of the measured endpoint, R is defined in Equation 3.1, $g$ is the natural logarithm of one of the three concentration-effect models, $\theta_g$ is the vector of its parameters and $\epsilon \sim \mathcal{N}(0, \sigma)$.

We performed a model comparison with the following ideas in mind:

- for the rest of the analysis, we wish to use the same model to describe the response of all the diatom species to a given contaminant.
- parsimony is a criterion to consider, as each additional parameter means additional complexity for the hierarchical modelling.

We fitted the three models on all (species, herbicide) pairs to judge how well they described the data (Figure 3.5 for atrazine only, Figure C.2 for all herbicides). The fit was performed using non linear least square regression[Bates and Watts, 1988] with the R[R Core Team, 2013] function *nls*. The first observation is that it was not possible to fit the BC and the CRS models for all (species, herbicide) pairs. The increased flexibility gained from the additional parameters rendered likelihood maximisation difficult. Several strategies for the starting values were tested, including that of using the estimates from R package *drc*[Ritz and Streibig, 2005] which is dedicated to fitting these models, but none permitted fitting all the (species, herbicide) pairs. The log-logistic model, however, could be fitted for all (species, herbicide) pairs. The second observation is that once log-transformed, the data did not show an important hormesis effect anymore. The three models tended to coincide in the region of large concentrations. In the region of small concentrations where the models differed, the BC and the CRS models did not appear to provide a significant improvement to the fit in general (except maybe for atrazine, NPAL). The case of atrazine, ADMI (Figure 3.5) shows that the CRS model might also suffer from poor robustness. The apparent hormesis seems to result from the influence of a single point. If removed, the CRS model would probably show only a very small hormesis effect.

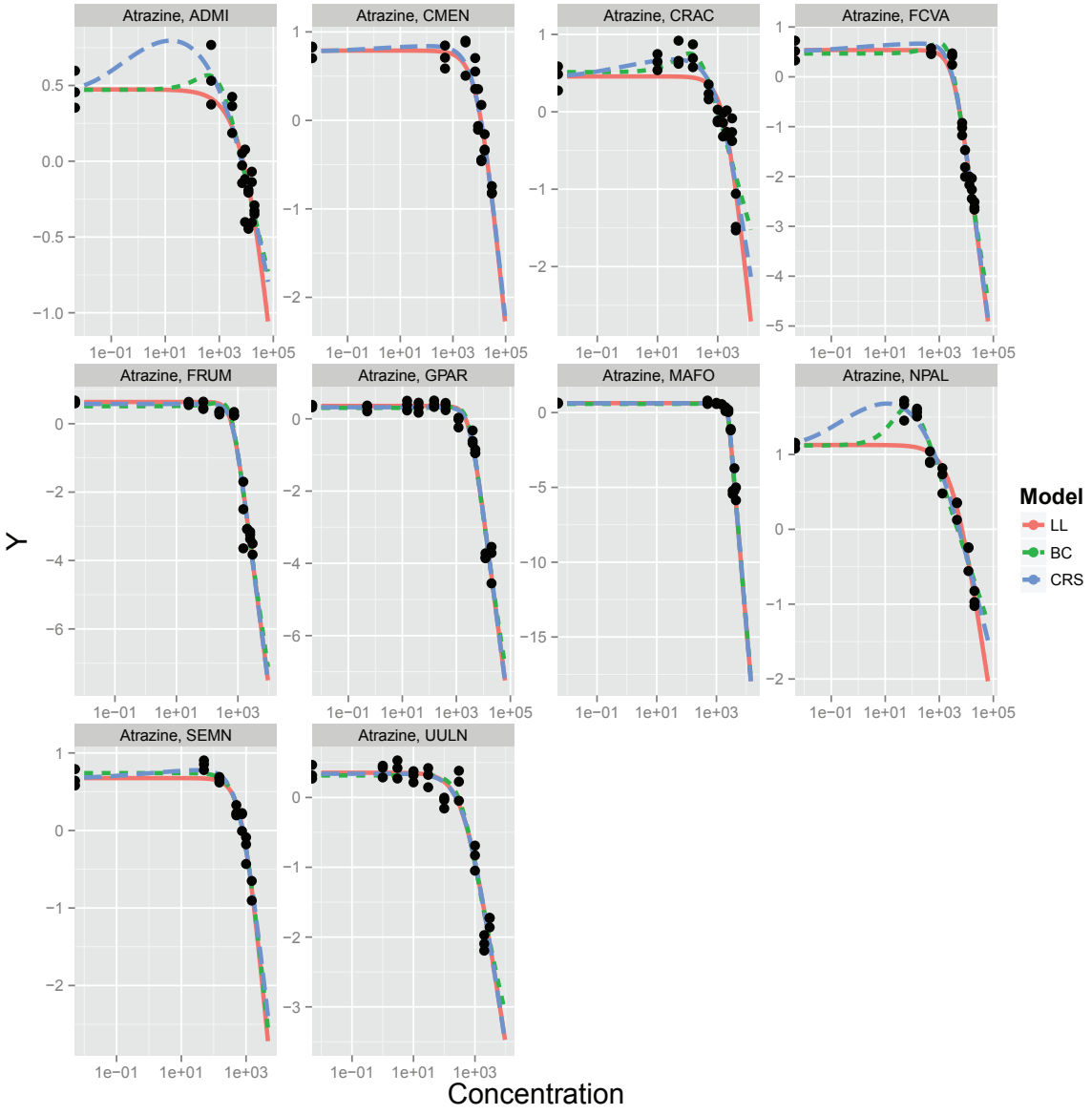



Figure 3.5: Comparison of the fit of the three models on the log-transformed data. LL stands for log-logistic, BC for Brain-Cousens and CRS for Cedergreen-Ritz-Streibig.

We also computed the AIC for each fit (detailed results in Appendix section C.3) to check if increasing model complexity really improved model fit. Burnham[Burnham et al., 2011] mentioned that early literature on AIC recommended that model 2 was to be preferred over model 1 if $AIC_1 - AIC_2 > 2$. He added that this rule was now known to be poor in general and that if $7 > AIC_1 - AIC_2 > 2$ or $-7 < AIC_2 - AIC_1 < -2$, model 1 should rarely be dismissed. To estimate the proportion of (species, herbicide) pairs for which a model more complex than the log-logistic offered a better description of the data we used the following heuristic rule:

- If a model cannot be fitted, then the log-logistic model is preferred.
- If AIC CRS − AIC LL $> -7$ or AIC BC − AIC LL $> -7$, then there is no strong evidence against the log-logistic model in favour of the more complex model, so following a parsimony principle we conclude that the log-logistic model is preferred. (this corresponds to an evidence ratio lower than 33.1 for the more complex model against the log-logistic model)[Burnham et al., 2011].

Using this heuristic rule, the log-logistic model was preferred over the CRS model for 94 % of the dataset and over the BC model for 87 % of the dataset.
If we made this rule less stringent on the AIC criterion, choosing for instance $\Delta\text{AIC} > -2$ (this corresponds to an evidence ratio lower than 2.7 for the more complex model against the log-logistic model), the log-logistic model was still preferred over the CRS model for 85 % of the dataset and over the BC model for 79 % of the dataset.
In summary, an hormesis effect seemed present in the original data, but a log-transformation to stabilise the variance made this effect look less important. Visually assessing the model fits lead to conclude that the more complex hormetic models did not provide a much better fit while sometimes posing convergence problems. Using AIC as a criterion for model comparison did not tip the scales in favour of one of the hormetic models against the log-logistic model either. Therefore, the log-logistic model was selected for use in the hierarchical model. Note however that in principle, the hierarchical approach could have been developed on either of these models.
The model retained was:

$$\mathrm{Y} = \ln(\mathrm{R}) = \ln\left(\frac{d}{1 + \left(\frac{C}{e}\right)^{b}}\right) + \epsilon \tag{3.6}$$

where $\epsilon \sim \mathcal{N}(0, \sigma)$.
In what follows, as parameter $d$ was estimated as the mean of the response in the control replicates for all the herbicides, parameters $b$ and $e$ were estimated by fitting the model from Equation 3.6 to data at the other concentrations, to avoid using data twice. This was not the case for the model comparison study because AIC for different models must be compared on the same dataset. Another reason to choose to estimate parameter $d$ separately was because we were not interested in modelling or predicting the response in the control experiment. Only parameters $b$ and $e$ characterise the effect of the herbicide on the diatom species.

## 3.3 Computing the classical Species Sensitivity Distribution

We wanted to compare our hierarchical SSD to a reference log-normal SSD. For each concentration-effect curve, we first fitted the model from Equation 3.6) by non-linear regression using the R function *nls* and the R package *nlstools*[Baty et al., 2015]. We extracted the $\mathrm{EC}_{10}$ and the $\mathrm{EC}_{50}$ in order to compare two levels of effect. We computed bootstrap 95% confidence intervals on the $\mathrm{EC}_{10}$ and $\mathrm{EC}_{50}$ using non-

parametric bootstrapping (results on Figure 3.6). Then, we fitted two log-normal distributions to the set of $EC_{10}$ and $EC_{50}$ using maximum likelihood via the web-tool MOSAIC_SSD[Kon Kam King et al., 2014] and obtained the $HC_5$ for the community, with a bootstrap 95% confidence interval as well. The result for $EC_{10}$ was $HC_{5,EC_{10}} = 1.3\ \mu g.l^{-1}$ [0.41; 5] and for $EC_{50}$ was $HC_{5,EC_{50}} = 38\ \mu g.l^{-1}$ [17; 94].

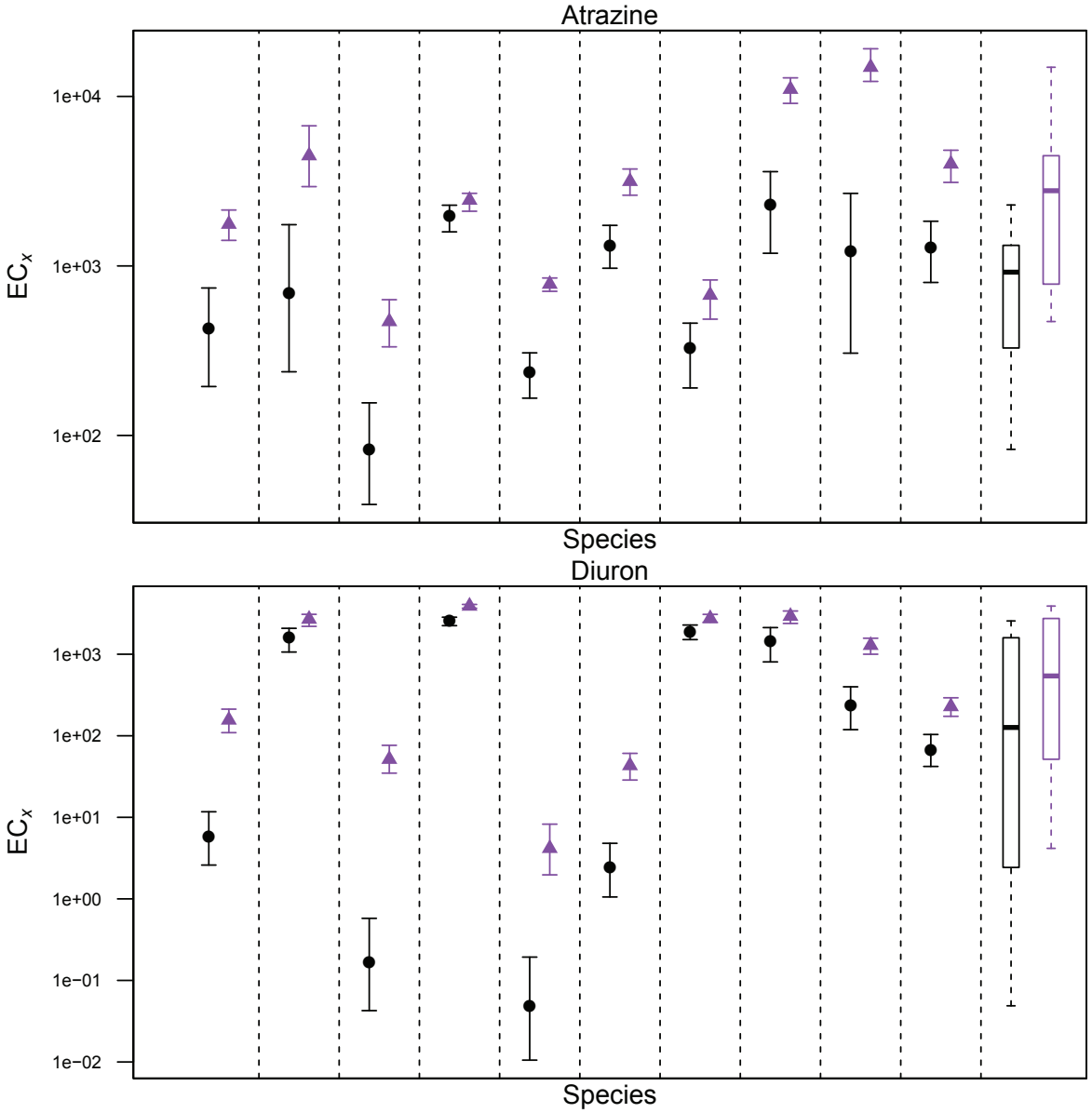



Figure 3.6: $EC_{10}$ in black and $EC_{50}$ in purple for each species, with the 95% bootstrap confidence interval. Vertical dotted lines separate each species. The box at the right hand side of each plot is the distribution of the point estimates of the $EC_x$ at the corresponding level of $x$ for all the species. $EC_x$ are in $\mu g.L^{-1}$.

Table 3.2: Description of the links indicated in Figure 3.7

| Node | Type | Equation |
|---|---|---|
| $(\log(\mathsf{b}_j), (\log(\mathsf{e}_j))$ | Stochastic | $(\log(\mathsf{b}_j), (\log(\mathsf{e}_j)) \sim \mathcal{N}_m(\mu, \mathbf{\Sigma})$ |
| $R_{i,j}$ | Deterministic | (Equation 3.2) |
| $Y_{i,j}$ | Stochastic | $\mathsf{Y}_{i,j} \sim \mathcal{N}(\ln(R_{i,j}), \sigma)$ |

$\mu = \begin{pmatrix} \mu_{\log b} \\ \mu_{\log e} \end{pmatrix}$, $\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_{\log b} & \rho\sigma_{\log b}\sigma_{\log e} \\ \rho\sigma_{\log b}\sigma_{\log e} & \sigma^2_{\log e} \end{pmatrix}$, $\mathcal{N}$ is the normal distribution and $\mathcal{N}_m$ is the multivariate normal distribution.

## 3.4 Hierarchical species sensitivity distribution

### 3.4.1 Structure of the model

A hierarchical approach is very different from the fitting of individual concentration-effect curves. The philosophy behind the hierarchy is that all tested species represent a random sample from the community and that their responses follow a distribution. More precisely, parameters $b$ and $e$ of the concentration-effect model are assumed to follow a multivariate distribution. This reasoning is an intuitive extension of the classical SSD, where CECs of the species are assumed to follow a community sensitivity distribution. A difference with the classical SSD approach is that the parameters of the community, called the hyperparameters, are estimated in one stroke from all the experimental data. This provides the advantage of pooling all the information together. Species for which the data are of very good quality will have the most important contribution to the global fit. Species for which the response is not characterized very precisely (large uncertainty on the parameters), or where data are missing, contribute less. In other words, all the data contribute to the estimation of the parameters at the extent of the information they contain. The classical SSD approach, on the contrary, heavily relies on the quality of the CEC estimates, and the requirements may be severe [Dowse et al., 2013]. In the previous study of the diatom dataset [Larras et al., 2012], this entailed discarding all the data on which it was not possible to fit a concentration-effect model.
Figure 3.7 sketches the hierarchy in the model and Table 3.2 describes the links of the model. Parameter $d$ having been estimated separately, we modelled the joint distribution of parameters $b$ and $e$. Both of them were assumed to follow a log-normal distribution. The log-normal distribution is the most commonly used distribution for parameter $e$ [Wheeler et al., 2002] (which corresponds to the $EC_{50}$). We assumed the same distribution for parameter $b$, knowing that the small number of species does not allow a better informed choice of distribution. There could be a correlation between these two parameters, which we also modelled (parameter $\rho$). Therefore, $\log(b)$ and $\log(e)$ were assumed to follow a multivariate normal distribution (log is used for the base-10 logarithm).

### 3.4.2 Bayesian methods

We used JAGS [Plummer, 2003] to fit the hierarchical model. JAGS performs Bayesian inference using Gibbs sampling via Markov Chain Monte Carlo (MCMC) simulation. The

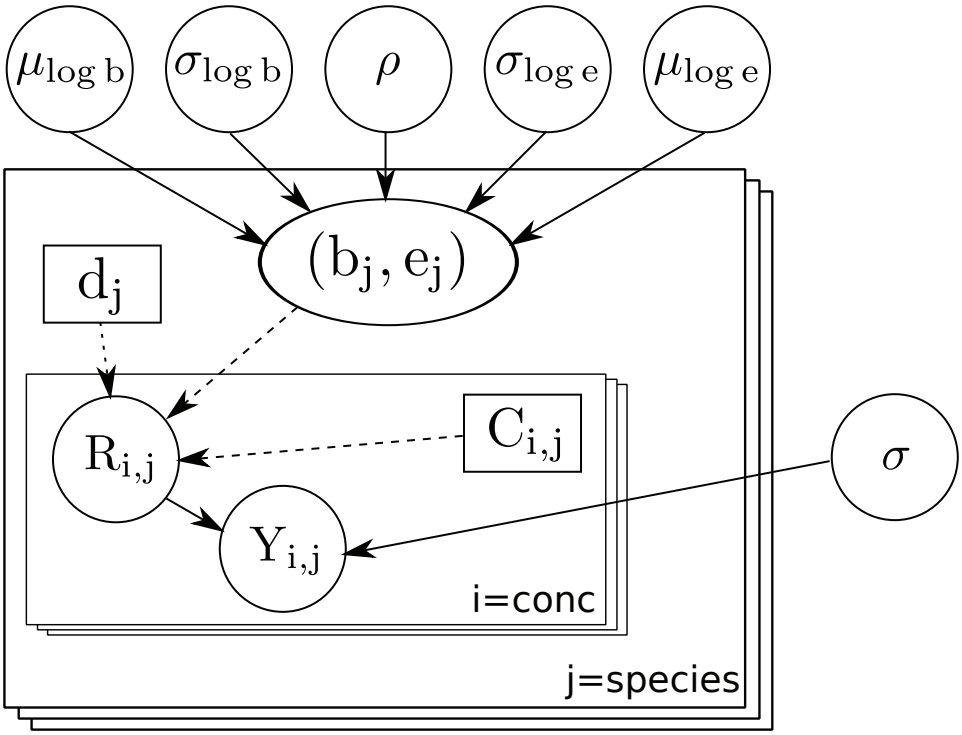



Figure 3.7: Probabilistic directed acyclic graphical model of the hierarchical model [Koller and Friedman, 2009](also called Bayesian network). Ellipses represent variables, rectangles represent covariables. Solid lines represent stochastic links, dotted lines represent deterministic links. To avoid repetition, similar subunit are summarized by plates. The inner set of plates denote the different concentrations, the outer set of plates denote the different species. Since the graph is directed and acyclic, any two variables are conditionally independent given the value of their parents.

Table 3.3: Prior distributions used for the hyperparameters of the hierarchical model (Figure 3.7)

| Parameter | Distribution | Type of prior |
|---|---|---|
| $\mu_{\log b}$ | $\sim \mathcal{U}(-6, 6)$ | vague |
| $\sigma_{\log b}$ | $\sim \mathcal{U}(0, 10)$ | vague |
| $\mu_{\log e}$ | $\sim \mathcal{N}(\mu_{\log \mathrm{C}}, \sigma_{\log \mathrm{C}})$ | informed by the concentration range |
| $\sigma_{\log e}$ | $\sim \mathcal{U}(0, 10)$ | vague |
| $\rho$ | $\sim \mathcal{U}(-1, 1)$ | vague |
| $\sigma$ | $\sim \mathcal{U}(0, 2)$ | vague |

$\mu_{\log \mathrm{C}} = \frac{\log(\min(\mathrm{C_{i,j}}))+\log(\max(\mathrm{C_{i,j}}))}{2}$ and $\sigma_{\log \mathrm{C}} = \frac{\log(\max(\mathrm{C_{i,j}}))-\log(\min(\mathrm{C_{i,j}}))}{4}$.
$\mathcal{N}(\mu, \sigma)$ denotes the normal gaussian distribution of mean $\mu$ and standard deviation $\sigma$, $\mathcal{U}(a, b)$ denotes the uniform distribution between $a$ and $b$.

priors are detailed in Table 3.3. The prior on $\mu_{\log e}$ was a normal distribution centred on the middle of the range of all tested concentrations. Its standard deviation was defined so that $\mu_{\log e}$ had a 95% probability to lie between the largest and the smallest tested concentrations. All the other priors were vague priors. The chains were run for 500 000 iterations and one in 40 were conserved.
The convergence of three chains was checked computing the Gelman-Rubin diagnostic [Brooks and Gelman, 1998]. Prior and posterior distributions were compared to check visually that the priors did not constrain the estimation of the posteriors. The relative width of the prior and posterior distributions was also compared to ensure that sufficient information was learned from the data. The parameters of the hierarchical model came out as a joint posterior distribution. The median of the marginal distributions were used as estimates of the parameters. The 2.5 and 97.5 percentiles of the distribution were used to define a 95% credible interval (the concept of credible interval is described in Appendix section B.5). The JAGS code to fit the hierarchical model is provided in Appendix D.

## 3.4.3 Results from the fit of the model

### 3.4.3.1 Convergence of the MCMC algorithm

The MCMC chains converged for all contaminants, according to the Gelman-Rubin statistics [Brooks and Gelman, 1998]. Figure 3.8 shows for diuron that except for parameter $\rho$, the vague prior distributions did not constrain the posterior distribution of the parameters and that there was sufficient information in the original data to estimate them. Similar results were observed for the other herbicides. The apparent constraint on correlation parameter $\rho$ is natural, since the correlation has to lie between 0 and 1. The fit of the model was visualised at the level of the diatom species by superimposing the fitted curves on the original data. The fitted curves were obtained by taking the median values of parameters $b_j$ and $e_j$ from the marginal posterior distributions. Figure 3.9 shows that for all herbicides, the estimation of the global parameters of the hierarchical model corresponds to a good fit at the level of the diatom species.

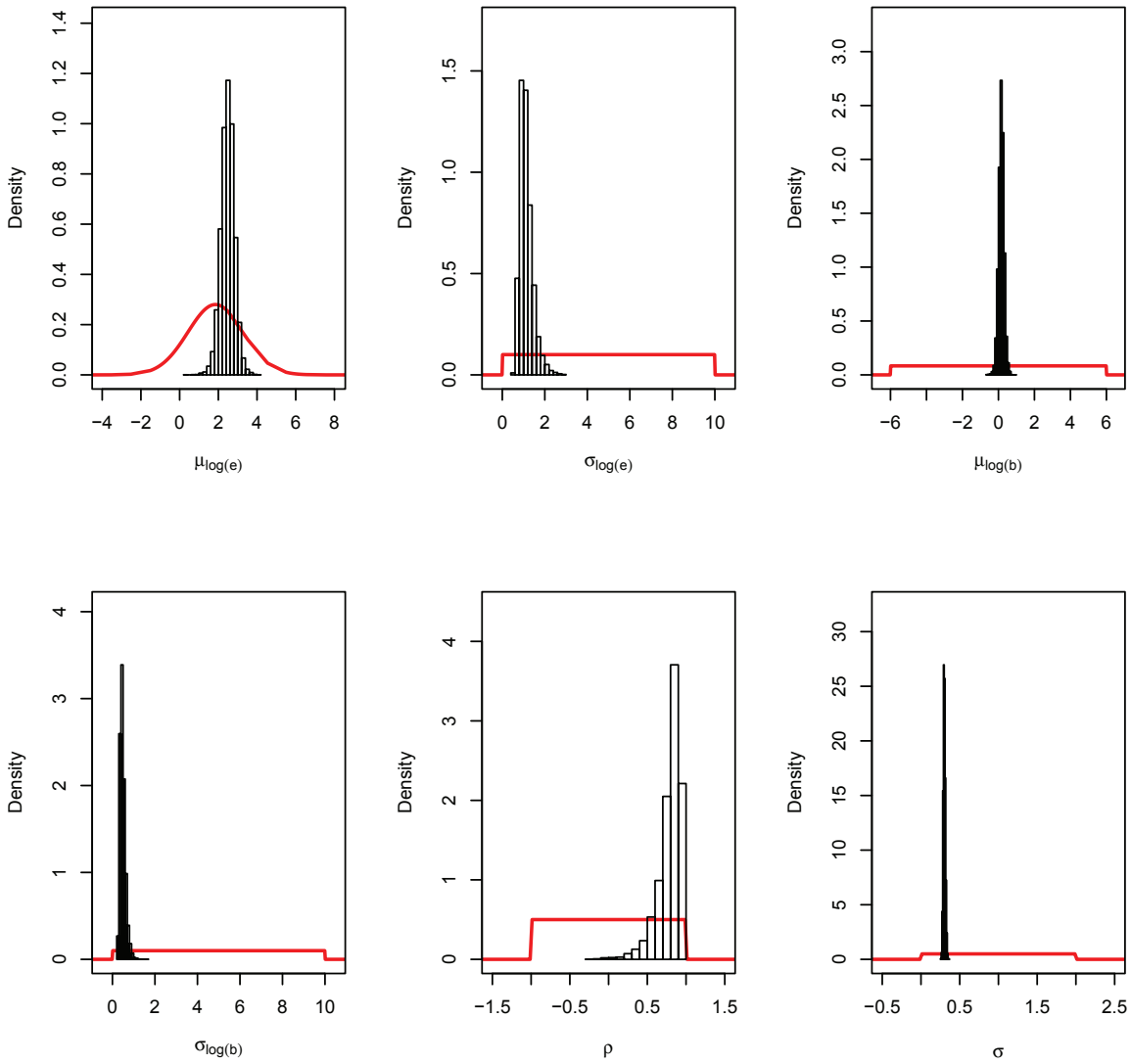



Figure 3.8: Comparison of the priors (in red) defined in Table 3.3 with the marginal posterior distributions (black histograms) for diuron.

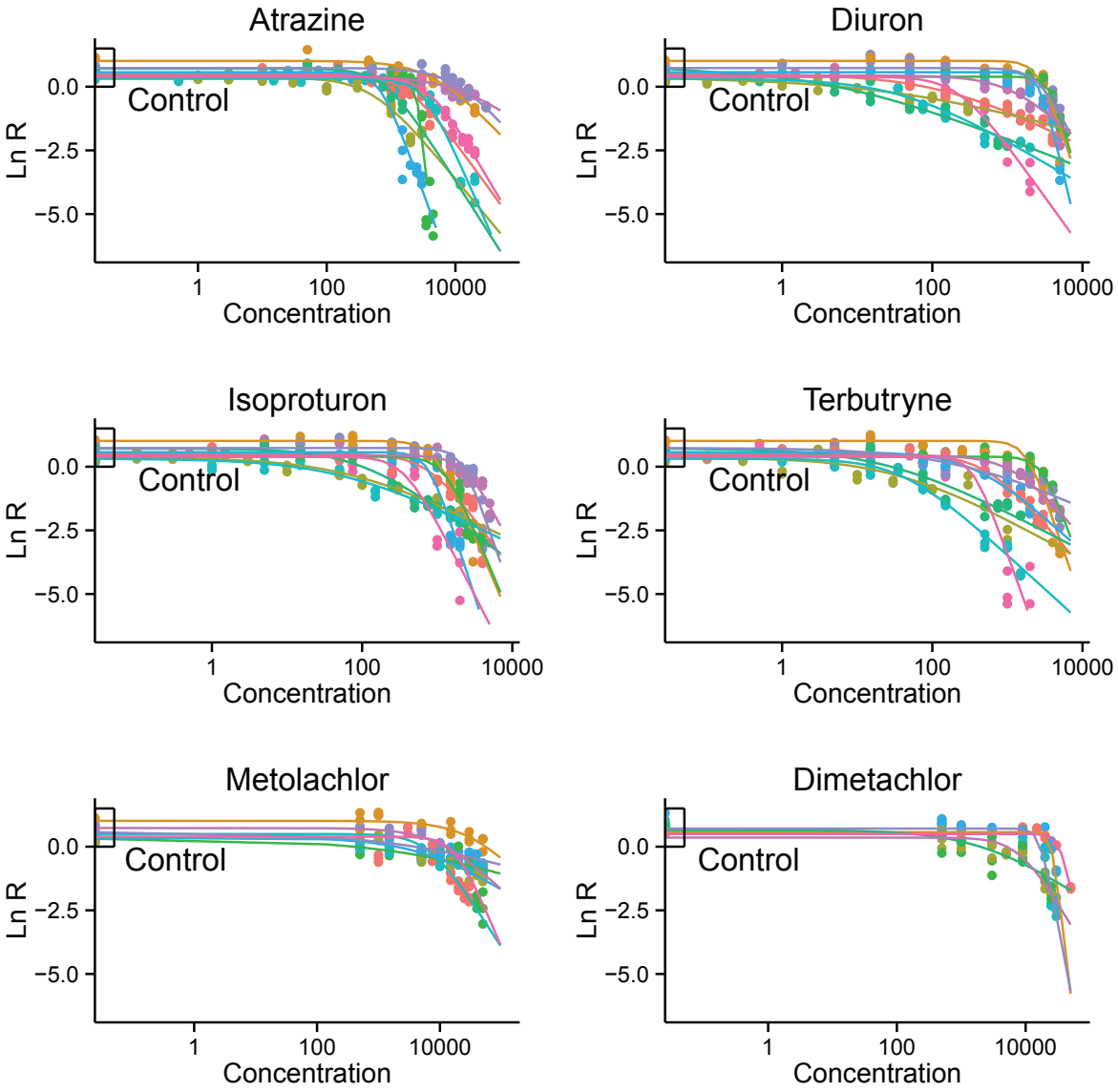



Figure 3.9: Original data and fit of the model at the level of the diatom species for all herbicides. Each colour denotes a different species. The rectangle denotes the control response in log scale. The curves are the median values of the posterior survival probabilities as a function of the concentration. Concentrations are in $\mu g.L^{-1}$.

#### 3.4.3.2 Estimated parameters

The estimated parameters for each herbicide are presented in Table 3.4, along with their 95% credible interval. For atrazine, the 95% credible interval on the correlation parameter $\rho$ was centred around 0 and did not suggest the presence of correlation between parameters $b$ and $e$. For all the other herbicides, however, there was a correlation between these two parameters. Slope parameter $b$ qualitatively determines how a species is affected by the contaminant: for a small value of $b$, the species is gradually affected by the contaminant, whereas for a large value of $b$, this species is almost insensitive to the

contaminant up to a certain threshold, then suffers a drastic effect. A strong positive correlation between $b$ and $e$, the slope parameter and the $EC_{50}$, implies that species with a low slope parameter also have a small $EC_{50}$, i.e. the most sensitive species are affected gradually. A positive correlation also implies that the most resilient species show no effect up to a certain threshold, followed by a sudden drop in fluorescence. In the absence of correlation, there is no constraint on the relative value of $b$ and $e$ for a given species and all types of behaviours can be encountered. The effects of that correlation are apparent on Figure 3.9: for atrazine, the contaminant for which species showed no correlation between parameters $b$ and $e$, exhibits all sorts of behaviours. For diuron, the species with small $EC_{50}$ have a gradual slope and those with large $EC_{50}$ have a steep slope. Such information about correlation between the concentration-effect parameters is not considered or taken advantage of using the classical SSD approach. Yet this is an information of biological relevance which can be addressed through the hierarchical modelling of SSD.
The robustness of the parameter estimation to changes of the vague prior distributions was assessed. The vague priors on all the parameters except the correlation parameter were shrunk by half and extended by half without a noticeable change in the posterior distributions. For the correlation parameter, a common practice is to give a prior distribution on the Fisher transform[3] of the correlation parameter [Gelman et al., 2013, Daniels and Kass, 1999] rather than on the correlation parameter itself. The uniform prior on the range $[-1; 1]$ was changed to a uniform distribution between $[-10; 10]$ and to a normal distribution $\mathcal{N}(0, 25)$ on the Fisher transformation of the correlation parameter. The resulting priors on the correlation parameter both had a density distribution concentrated around -1 and 1 and very little weight around 0, which favoured large correlation parameters (Figure 3.10). For an herbicide with a strong correlation such as terbutryn, there was no strong influence of the prior, as could be expected. For atrazine, which showed no strong correlation, we did not find a strong influence of the prior either.

## 3.4.4 Simulating from the model

### 3.4.4.1 Building a global response

The hierarchical SSD approach extracts more information from the raw data than the classical SSD. In particular, we can use it to compute an indicator of the global response of the community, which is a relevant information for the protection of that community. Once the model fitted, the joint posterior distribution of the parameters contained all the information that can be extracted from the data about the response of the community to the contaminant. For a set of global parameters $\boldsymbol{\theta} = (\mu_b, \sigma_b, \mu_e, \sigma_e, \rho)$ obtained from the posterior distribution, it was possible to reconstruct a full community by sampling individual species and to predict its response to the contaminant. Sampling a species $j$

[3]The Fisher transform of $\rho$ is defined as:

$$z = \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right) \tag{3.7}$$

Table 3.4: Median parameters of the hierarchical model and their 95% credible interval for each herbicide, along with the number of species tested for each contaminant. The median of the marginal posterior distribution is taken as the parameter estimate.

| | atrazine | diuron | isoproturon |
|---|---|---|---|
| **Parameter** | **Estimate** | **Estimate** | **Estimate** |
| $\mu_{\log b}$ | 0.28[0.02; 0.55] | 0.16[-0.15; 0.46] | 0.27[0.027; 0.48] |
| $\sigma_{\log b}$ | 0.37[0.22; 0.69] | 0.46[0.3; 0.82] | 0.33[0.21; 0.59] |
| $\mu_{\log e}$ | 3.4[2.9; 3.8] | 2.5[1.8; 3.2] | 2.6[2.2; 3.1] |
| $\sigma_{\log e}$ | 0.58[0.37; 1.1] | 1.1[0.7; 1.9] | 0.66[0.42; 1.2] |
| $\sigma$ | 0.38[0.35; 0.43] | 0.3[0.28; 0.33] | 0.39[0.36; 0.43] |
| $\rho$ | -0.22[-0.74; 0.47] | 0.83[0.39; 0.96] | 0.86[0.47; 0.97] |
| Number of species | 10 | 10 | 8 |
| | terbutryn | metolachlor | dimetachlor |
| **Parameter** | **Estimate** | **Estimate** | **Estimate** |
| $\mu_{\log b}$ | 0.16[-0.089; 0.39] | -0.064[-0.45; 0.29] | 0.42[-0.31; 0.96] |
| $\sigma_{\log b}$ | 0.35[0.22; 0.65] | 0.4[0.22; 0.94] | 0.59[0.28; 1.5] |
| $\mu_{\log e}$ | 2.4[1.8; 3] | 4[3.6; 4.3] | 4[ 3; 4.5] |
| $\sigma_{\log e}$ | 0.83[0.52; 1.5] | 0.38[0.19; 0.97] | 0.61[0.29; 1.8] |
| $\sigma$ | 0.41[0.38; 0.46] | 0.32[0.28; 0.36] | 0.4[0.34; 0.47] |
| $\rho$ | 0.68[0.083; 0.92] | 0.42[-0.49; 0.9] | 0.85[-0.17; 0.99] |
| Number of species | 10 | 10 | 5 |

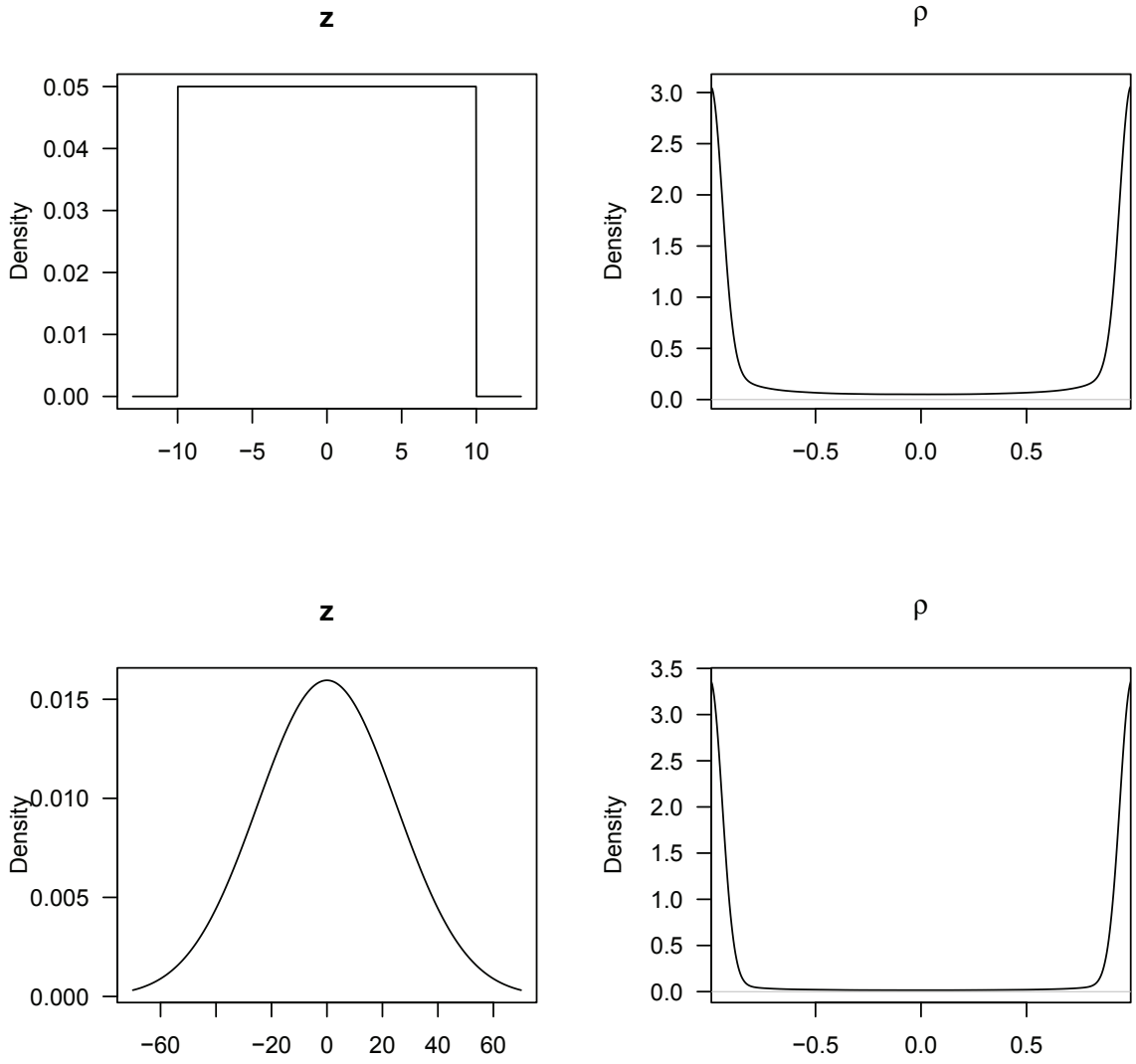



Figure 3.10: Prior on the correlation parameter $\rho$ implied by putting a prior on its Fisher transform $z$. Top shows $z \sim \mathcal{U}(-10, 10)$, bottom shows $z \sim \mathcal{N}(0, 25)$.

is equivalent to sampling a pair of parameters $(b_j, e_j)$ from the multivariate log-normal distribution parametrised by $\boldsymbol{\theta}$.
In order to predict the response of a realistic community, we chose to focus on finite-size communities. Diatom communities may number around 30 different species. Note that a specific draw of 30 species produces a community with a certain response and that another draw of 30 species would produce a different response. Therefore, there is some uncertainty in the response obtained for a group of 30 species, even assuming that $\boldsymbol{\theta}$ is known. Moreover, the $\boldsymbol{\theta}$ parameters themselves are uncertain and follow a distribution. These two sources of uncertainty were taken into account by sampling around 10 000 sets $\boldsymbol{\theta}_k$, then sampling 30 species for each $\boldsymbol{\theta}_k$.
After a community was simulated, we defined its global response as the global fluorescence of the community, depending on the concentration. The global fluorescence was defined as the sum of the fluorescence of each species.
To obtain a global response, we assumed that all species participated equally in the global fluorescence. Following this assumption, it was possible to define an indicator of the global response of the community at a given concentration, called $\mathrm{r_{tot}}$:

$$\mathrm{r_{tot}} = \frac{\sum_{j \in species} \frac{\mathrm{R}_j}{\mathrm{R}_j^0}}{\mathrm{N_{species}}} \tag{3.8}$$

where $\mathrm{R}_i$ is the response of species $i$ at a given concentration, and $\mathrm{R}_i^0$ the response in the control experiment. The indicator $\mathrm{r_{tot}}$ of the global response is a quantity between 0 and 1 which describes the global reduction in fluorescence growth compared to the control, as a function of the concentration in contaminant. Analogous to the $\mathrm{HC}_5$ for the SSD, a Global Effect Concentration at 5% ($\mathrm{GEC}_5$) was defined, which corresponds to the concentration leading to a reduction of 5% of the global response $\mathrm{r_{tot}}$. In our case, the $\mathrm{GEC}_5$ corresponds to a reduction of 5% of the community fluorescence ($\mathrm{r_{tot}} = 0.95$). Following the terminology used for SSD in Posthuma [Posthuma et al., 2010], the hierarchical SSD, and more precisely the prediction of the global response, can be used in an *inverse* and a *forward* manner. The *inverse* approach consists in setting a protective concentration threshold, the $\mathrm{GEC}_5$, below which 95% of the global response of the community should be protected. The *forward* approach consist in determining the reduction in the global response of the community for a given concentration level.
Figure 3.11 shows the importance of considering the global response of the community for risk assessment. The top of Figure 3.11 shows the $\mathrm{HC}_5$ obtained using the classical SSD approach on the $\mathrm{EC}_{10}$ endpoint, while the middle shows the $\mathrm{HC}_5$ obtained using the $\mathrm{EC}_{50}$. This concentration is used for regulatory purposes as the Predicted No Effect Concentration (PNEC), which determines the threshold under which the community is considered protected. The $\mathrm{HC}_5$ only aims at preventing a proportion of the species from being harmed, disregarding the possibility that harming key species could endanger the whole community. In order to protect the community in terms of the endpoint measured in the original data (fluorescence, biomass), it is interesting to consider also the $\mathrm{GEC}_5$ in the risk assessment. In the case of atrazine, the concentration which induces a reduction of 5% of the global biomass ($\mathrm{GEC}_5$) is lower than both the $\mathrm{HC}_5$ based on the $\mathrm{EC}_{10}$

and on the $EC_{50}$ (Figure 3.11). In the case of diuron, the $GEC_5$ is much lower than the $HC_5$ based on the $EC_{50}$ and similar to the $HC_5$ based on the $EC_{10}$. For the four other herbicides, the $GEC_5$ is between the two $HC_5$ and in general, the $GEC_5$ is close to the $HC_5$ based on $EC_{10}$. Calculating the reduction in global biomass at the $HC_5$ (i.e. using the *forward* SSD approach) indicates that for atrazine, the classical $HC_5$ built on the $EC_{50}$, which protects 95% of the species, could protect only 81% [55%,94%] of the global biomass. The classical $HC_5$ based on the $EC_{10}$ could protect only 92% [73%,99%] of the global biomass. In the case of diuron, the classical $HC_5$ built on the $EC_{50}$ protects $86\%[69\%, 96\%]$ of the biomass, while the classical $HC_5$ built on the $EC_{10}$ protects $96\%[86\%, 100\%]$ of the global biomass.
To summarize the comparison, there is no systematic relationship between the $GEC_5$ and the $HC_5$. Aiming to protect 95% of the global response of the community could prove either more or less protective than aiming to protect 95% of the species. But it is important to note that for atrazine and diuron, a $HC_5$ based on the $EC_{50}$ might protect only $80 - 86\%$ of the global response of the community.

#### 3.4.4.2 Species Sensitivity Distribution as a function of the level of effect ($x$ of the $EC_x$)

We have considered a new indicator, the global response of the community. From the fitted model, it is also possible to reconstruct an SSD. In this case, the simulation aimed at representing the variability of species sensitivity, i.e. the distribution of any $EC_x$ in the community. For 2000 sets of global parameters $\boldsymbol{\theta}_k$, the effects (Equation 3.2) for each species of a community were simulated, and their $EC_x$ calculated. Estimating an $HC_5$ for a community consists in determining the fifth percentile of that $EC_x$ distribution. To get the best estimation of the fifth percentile, large communities were simulated ($4 \times 10^6$ species). An SSD (and $HC_5$) with a 95% credible interval was estimated for these 2000 sets of parameters, from the median, $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles. This SSD is an improvement on the classical SSD, since it was estimated taking into account all the information present in the original data and accounting for the potential inter-species correlation among the parameters b and e of the concentration-effect model. In particular, the uncertainty in the estimation of the parameters of the concentration-effect curves was propagated in the SSD estimation. Another advantage of reconstructing the SSD from the fitted hierarchical model is that it does not require to choose the $x$ of the $EC_x$ in advance, contrary to the classical SSD, which starts with a certain level of $EC_x$. Fitting a classical SSD on another level of effect requires going back to the original data. Using the fitted model and the simulation scheme instead, it is possible to calculate an SSD on any $x$ of the $EC_x$. We used our hierarchical model to study how the $HC_5$ may vary as a function of the $x$ of the $EC_x$.
Figure 3.12 shows the $HC_5$ as a function of the $x$ of the $EC_x$ for the diatom community exposed to diuron, computed from the hierarchical SSD. This is an $HC_5$ which includes the uncertainty from the original raw data. The prediction from the hierarchical model was compared to the prediction from the classical SSD. The first striking observation is that the classical $HC_5$ based on the $EC_{10}$, which ignores the uncertainty from the determination of the CECs, is much higher than the hierarchical $HC_5$. The second observation is that for a hierarchical $HC_5$ which includes the original uncertainty on the

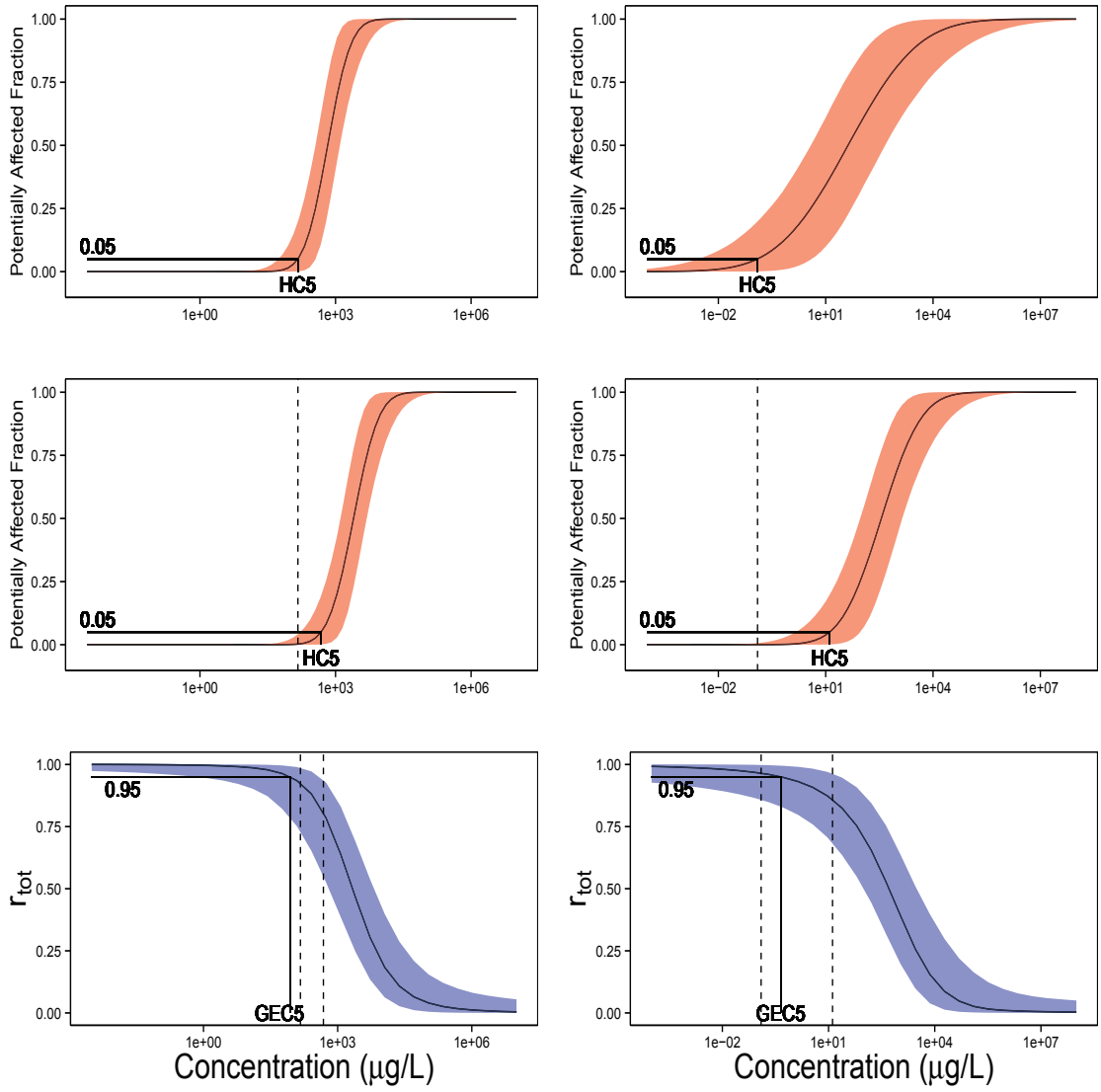



Figure 3.11: Species sensitivity distribution and global response of the community for atrazine (left) and diuron (right). Top: classical SSD built on the $EC_{10}$ with 95% bootstrap confidence intervals and the $HC_5$. Middle: classical SSD built on the $EC_{50}$ with bootstrap confidence intervals and the $HC_5$. Bottom: global response of the community with 95% credible confidence intervals and the concentration corresponding to a reduction of 5% of the global response ($GEC_5$). The vertical dotted lines are intended as visual cues for comparing the $HC_{5,EC_{10}}$, $HC_{5,EC_{50}}$ and the $GEC_5$.

CECs, the confidence intervals expand wildly for an $x$ of the $EC_x$ below 50. This can be linked to the fact that for small values of $x$, the uncertainty on the $EC_x$ estimated from a concentration-effect curve is larger than on the $EC_{50}$ (as was mentioned in the results for classical SSD for the $EC_{10}$, see also Figure 3.6). Therefore, in estimating the $HC_5$, the effect of discarding the uncertainty should be greater. This phenomenon cannot be observed with classical SSD, since it does not take into account uncertainty on the CECs. Such an observation contrasts with the reasoning from Aldenberg and Rorije [Aldenberg and Rorije, 2013], which stated that taking uncertainty on the CECs into account should increase the value of the $HC_5$. However valid, their argument cannot be directly applied to our case, for it rests strongly on the assumption of log-normality of the CECs, be they NOEC, QSAR or $EC_x$ at any level of effect. In our model, the $EC_{50}$ are assumed to follow a log-normal distribution (parameter $b$ and $e$ follow a log-normal distribution), which implies that $\log \mathrm{EC_x} = \log e + \frac{1}{b} \log\left(\frac{x}{1-x}\right)$ for any other $x$ than 50 does not follow a normal distribution. Therefore, it is not surprising to find a hierarchical $HC_5$ different from the classical $HC_5$.

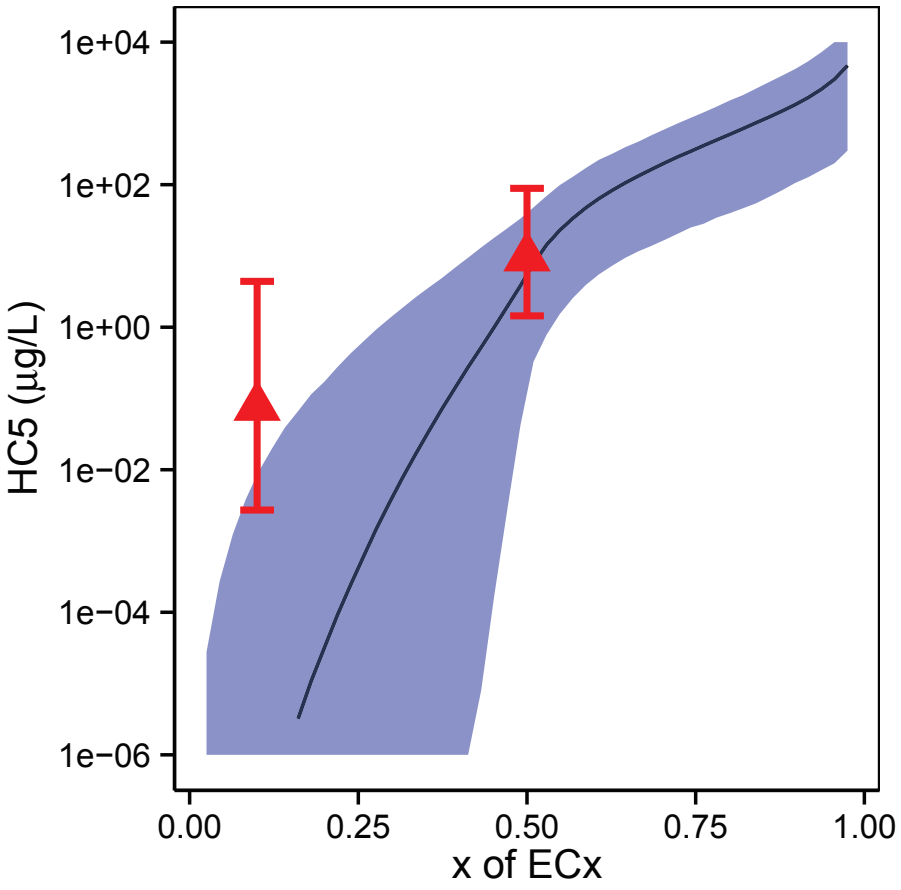
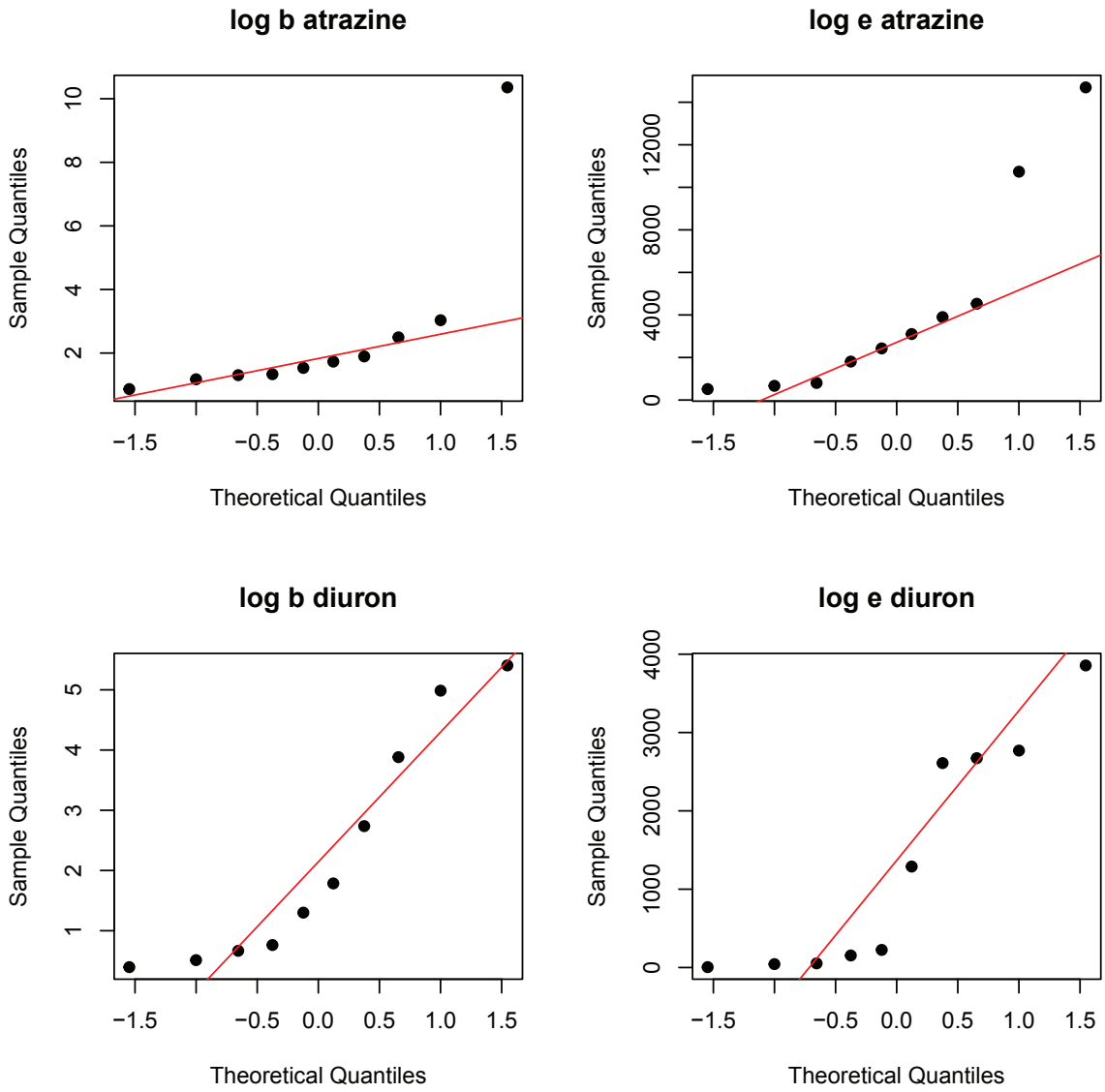


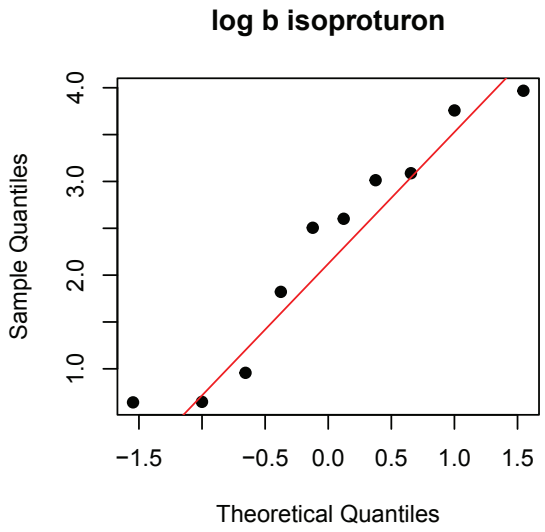

Figure 3.12: $HC_5$ as a function of the $x$ of the $EC_x$ for diuron obtained from the hierarchical SSD, with the 95% credible bands. In red, the $HC_5$ obtained from the classical SSD based on the $EC_{10}$ and the $EC_{50}$ with bootstrap confidence intervals.

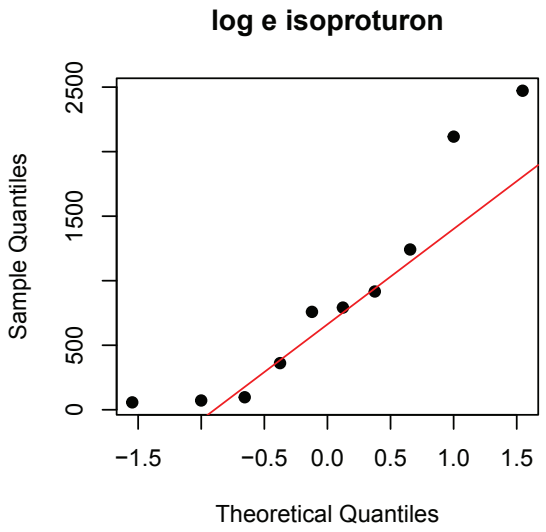
### 3.4.5 Sensitivity to modelling choices

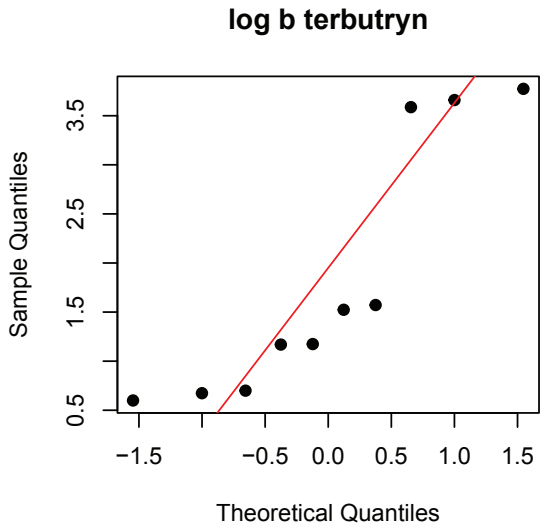
The first advantage of the hierarchical approach was to introduce more ecological relevance to the risk assessment than the bare classical SSD by taking into account all the biological information available in the data. The hierarchical approach also provides a perspective on the treatment of uncertainty in the classical SSD. Classical SSD adopts the same approach whatever the level of effect chosen. Yet, the degree of uncertainty can strongly depend on the level of effect and neglecting that uncertainty might certainly bias the estimation of the $HC_5$. The hierarchical SSD, which correctly propagates the uncertainty from the raw data to the $HC_5$ and builds the SSD on any $EC_x$, does not suffer from this problem. In particular, the hierarchical SSD shows that building an SSD on $EC_{10}$ without considering the uncertainty on these $EC_{10}$ might lead to a wrong estimate of the $HC_5$ and of its confidence interval. To obtain this result, we simply assumed that parameter $e$ followed the usual log-normal distribution [Wheeler et al., 2002] and opted for the same distribution law for the $b$ parameter. With at most ten species per contaminant, there is not much ground to argue for other distributions, but in the future it would be very interesting to analyse larger datasets. More tested species would provide a better characterisation of the distribution laws for the concentration-effect model parameters and might support the current distribution choice or guide towards a different structure for the hierarchical model. At any rate, quantile-quantile plots of the species parameters did not show a strong deviation from log-normality (Figure 3.13).
In order to explore this issue, we tried to use a multivariate log-t distribution instead of a multivariate log-normal distribution for the distribution of the parameters of the concentration-effect model in the community. The t distribution with location and scale parameters has one more parameter than the normal distribution, $\nu$, which can be used to tune the weight of the tails. With a large $\nu$, the t distribution is close to the normal distribution. The weight of the tails increases with decreasing $\nu$. We first tried to estimate this parameter on the data, giving it uniform priors of varying lengths, but there was not enough information in the data to estimate $\nu$ (the posterior distribution coincided with the prior). Given that there are at most 10 points per herbicide, this is not so surprising. Then, we checked the robustness of the results to increasing tail weight. We fitted hierarchical models with $\nu = 100, 50, 25, 12, 6, 3$ and 1 which corresponds to distributions with increasingly fat tail and compared the posterior distributions, the global response and the value of the $HC_5$ as a function of the level of effect. The posterior distributions, notably the correlation parameter, were found to be very similar for all values of $\nu$. The global response was also quite insensitive to changes in $\nu$ except for the extreme case $\nu = 1$, for which the median global response was lower and the credible intervals larger. This effect was strong for atrazine and terbutryn, intermediate for diuron and isoproturon, and smaller for metolachlor and dimethachlor. The value of the $HC_5$ as a function of the level of effect and the bounds of the 95% credible interval showed no change down to $\nu = 6$, a small change for $\nu = 3$ and were strongly affected at $\nu = 1$. The lower bound of the 95% credible interval was always the most affected. This can be understood because adding weight to the tails of the distribution increases the probability of finding extremely sensitive species and stretches the span of the confidence interval. Overall, the results proved rather stable except for very heavy

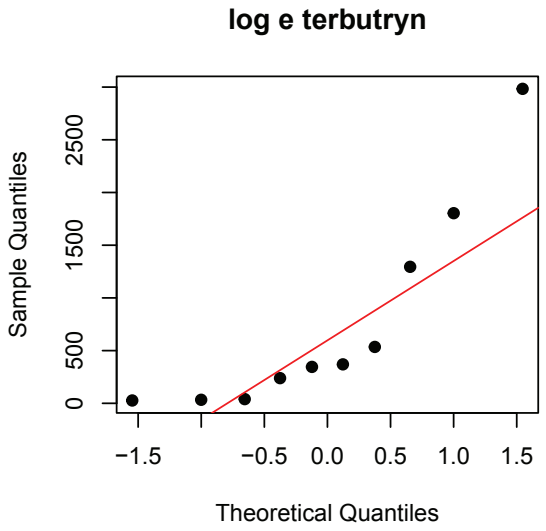
tailed distributions.




log b isoproturon
Sample Quantiles
Theoretical Quantiles
1.0
2.0
3.0
4.0
−1.5
−0.5
0.0
0.5
1.0
1.5


log e isoproturon
Sample Quantiles
Theoretical Quantiles
0
500
1500
2500
−1.5
−0.5
0.0
0.5
1.0
1.5


log b terbutryn
Sample Quantiles
Theoretical Quantiles
0.5
1.5
2.5
3.5
−1.5
−0.5
0.0
0.5
1.0
1.5


log e terbutryn
Sample Quantiles
Theoretical Quantiles
0
500
1500
2500
−1.5
−0.5
0.0
0.5
1.0
1.5

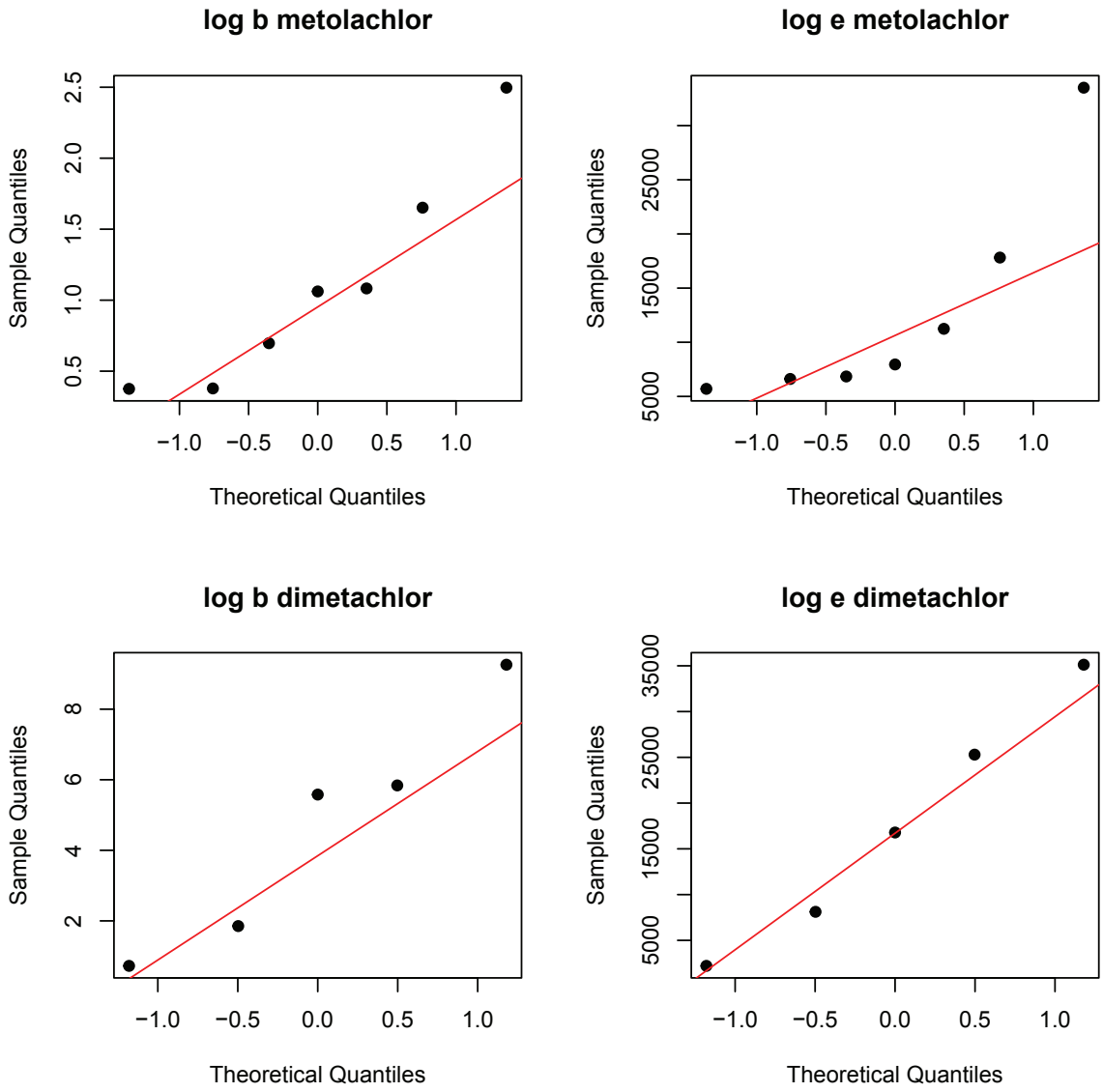



Figure 3.13: Q-Q plot of the random effects of the hierarchical model for each herbicide.

We mentioned earlier that our result on the hierarchical SSD including more sources of uncertainty and variability than the traditional analysis contrasted with the argument put forward by Aldenberg and Rorije in [Aldenberg and Rorije, 2013], where they explained that taking uncertainty into account should increase the estimate of the $HC_5$ compared to a classical SSD approach. We gave a first reason why their argument was not in contradiction with our work: in our model the $EC_x$s for $x$ different from 50 do not follow a log-normal distribution. A second reason is that in the hierarchical approach in [Aldenberg and Rorije, 2013], it is assumed that the uncertainty on the CEC (the length of the 95% confidence interval on the CEC) is identical for all species, whereas in our model the uncertainty on the parameters of the concentration-effect model is specific to each species. The species-specific uncertainty mostly depends on the quality of the raw

data for that species. This is important because the reasoning of Aldenberg and Rorije focuses on the estimation of the variance of the SSD, while including varying degrees of uncertainty for each species could affect the estimation of the mean of the SSD. To understand the role of varying levels of uncertainty, let us consider an extreme case: if there is a large uncertainty on the most sensitive species and a rather small uncertainty on all the other species, we can expect that taking uncertainty into account will shift the estimate of the mean upwards. Since the value of the $HC_5$ is a function of both the mean and the variance of the SSD, taking uncertainty into account, although reducing the variance of the SSD, does not necessarily increase the estimate of the $HC_5$. A third reason stems from the hierarchical structure of the model and the fact that the fit of the concentration-effect models at the level of the species is performed in one stroke. The fit of the $b$ and $e$ parameters for one species is influenced by the data from the other species. More specifically, since the tested species are assumed to come from the same community with the same species sensitivity distribution, the estimation of the concentration-effect model parameters is the result of information coming from all the species together. On the contrary, in classical SSD the fit of the concentration-effect parameters obtained by non-linear regression depends solely on the data for one species. Therefore, the $e$ parameter for a given species estimated in the hierarchical model can be slightly different from the $e$ parameter estimated by non-linear regression. Translated at the community level, this implies that the value of the $HC_5$ is not determined by the value of the CECs and their uncertainty, but by a more subtle interplay between the raw data and the distribution law of species sensitivity in the community. This is an example of the shrinkage property of a hierarchical model. For all these reasons, we do not believe that our results are incompatible with previous work by Aldenberg and Rorije[4] [Aldenberg and Rorije, 2013].

### 3.4.6 Discussion

Classical SSDs are widely used to assess risk of chemicals for natural communities, but they present certain limitations [Forbes and Calow, 2002, Power and McCarty, 1997]. In this part of the thesis, we presented a hierarchical approach to SSD, which includes all the information present in the raw bioassay data to overcome some of these limitations. This hierarchical SSD differs from classical SSD in that the whole concentration-effect curve is used to build the SSD instead of a single CEC per species. This implies that the hierarchical model requires the full output from bioassay response curves. Unfortunately, such data are not always available. For the three parameter log-logistic model used in this study, providing two CECs, such as the $EC_{10}$ and the $EC_{50}$ would be sufficient to describe the effect of the contaminant on a species. Therefore, reporting only two CECs at the end of a bioassay would be enough to construct a hierarchical SSD in the same spirit as that developed in this work, though without propagating the uncertainty on the CECs.

Making full use of the bioassay data, the hierarchical SSD propagates not only all the uncertainty from the concentration-effect curve, but also all the information on the shape

[4]Fox showed very recently that there was a mistake in the reasoning of Aldenberg and Rorije[Fox, 2015a]

of the curve. It also unveils possible interspecies correlation among the parameters, which have a biological significance.
$HC_5$ thresholds obtained from our hierarchical approach were close to those previously obtained with a classical method on the same dataset [Larras et al., 2012]. Photosynthesis inhibitor herbicides (especially diuron, terbutryn and isoproturon) were clearly found more hazardous than chloroacetamide family herbicides. Photosynthesis inhibitors are known to exert a strong pressure on diatom communities especially because they prevent photosynthesis, which is a vital process for microalgae.
One of the advantages of the hierarchical approach is the prediction of the global response of a community as a concentration-effect curve which in turn makes it possible to derive a global effect concentration of $x\%$, the $GEC_x$. This new kind of threshold does not provide *a priori* information at the species level (what and how much specific species are affected) but it is a tool to make *a priori* risk assessment at the community level (response of all the species together). This appeared especially interesting for microbial community, for which chemical effects are often observed and reported at the community level for many endpoints (i.e. biomass [Coutris et al., 2011], respiration, photosynthesis, fluorescence, enzyme activity...). This global response does not require the choice of an arbitrary effect level such as the $EC_{50}$ or the $EC_{10}$. Moreover, on the tested contaminants, the hierarchical approach resulted in safe concentration levels which were very close to the classical $HC_5$ defined on the $EC_{10}$. This led us to think that from an operational point of view, the use of the global response should prove as protective as the classical SSD approach. The global response may also be used to provide structural or functional information depending on the structural or functional nature of the measured endpoint. In that last case, the global response would provide information on the functional response of a community and solve one of the problems in the SSD approach [De Laender et al., 2008, Kefford et al., 2012b]. Fundamentally, the global response is an indicator containing a radically different type of information compared to SSD. The $HC_5$ aims to protect 95% of the species in a community, but there is considerable uncertainty about the fate of the community if the 5% affected play a key role for some other properties of the community (such as the global response). The $GEC_5$ protects 95% of the global response, but does not say what proportion of the species are significantly affected (above a given level of effect). Together, both SSD and global response provide complementary means to assess the effect of a contaminant on a community. Both ought to be considered when defining acceptable levels of concentration for a contaminant.
Our definition of the global response strongly depends on the assumption of equipartition of species contribution to the global response. In communities of diatoms, one or several species may dominate and their contribution to the global biomass could be preponderant. However, it has been observed that the dominance and the diversity of diatom species within a community change across the seasons [Singh et al., 2010, Pesce et al., 2009]. Therefore, when considering the biomass over a year, the contribution of many species might be averaged, rendering our assumption of equirepartition more plausible. At any rate, this assumption is already present in the classical SSD approach [Forbes and Calow, 2002]. As the simulated species are unidentified, it is not possible to attribute a weight to each of them to sum their biomass. To circumvent this assumption, it could be possible to define groups of species having comparable biomass

and define weights according to these groups. On a larger dataset, it would certainly be interesting to adopt this approach.
To summarize, benefits of the hierarchical SSD approach include that (i) all the experimental data is taken into account, (ii) uncertainty from the concentration-effect model can be included in SSD, (iii) there is no need for an arbitrary preliminary choice for the level of effect, (iv) the hierarchical structure means that we address the issue of noise in individual studies by shrinkage. Disadvantages include that (i) there are more distributional choices to be made by the modeller and (ii) that the dependence of the final inferences on the data and model choices becomes less clear as complexity of the model increases.
Extension of this work could take two different directions. First, it would be natural to consider a supplementary level of hierarchy to model inter-herbicide variability. This is desirable because it would open the door to an across-chemical extrapolation. However, this extension would require an assumption similar to those of classical SSD, namely that the herbicides tested represent a random and representative sample of all the possible herbicides. This is conceptually more difficult than the classical species representativity assumption of SSD. Moreover, this new degree of hierarchy would require many distributional choices for the $\mu_{\log e,herbicide}$, $\mu_{\log b,herbicide}$ etc. Whereas there is a strong tradition of choosing a log-normal distribution for the SSD, there is no guide for choosing the distribution of the parameters at the herbicide level. In developing the *hSSD* software [Craig, 2013], Peter Craig has tackled this question and assumed a normal distribution for the equivalent of $\mu_{\log e,herbicide}$ and a gamma distribution for $1/\sigma^2_{\log e,herbicide}$. The *hSSD* software relies on a hierarchical approach to SSD complementary to that presented in this part of the thesis. Instead of including all the information from the raw data, *hSSD* includes exterior taxonomic and across-chemical information in the SSD to predict a scenario-based $\mathrm{HC}_p$ for a community whose species are all identified. The second possible extension of our work stems from the following observation: our hierarchical SSD is meant to avoid summarizing the full concentration-effect curve by a single critical effect concentration and to make the most of the available data. However, only data at the end of the experiment were used. Bioassay data often include a tracking over time of the contaminant effect and this information could be included as well in the SSD. Modelling time dependence would essentially consist in adding a supplementary level to the hierarchy. Studying the time component of SSD is particularly interesting because toxicity of a contaminant clearly evolves over time, yet the observation period is often constrained by practical considerations [Fox and Billoir, 2013]. The next part of the thesis will focus on including time dependence into the SSD approach to improve the accuracy and the biological relevance of its predictions.

Chapter 4

# Hierarchical modelling of time-resolved survival data

Toxicity data obtained from bioassay are often time-resolved. Indeed, experiments need to be carried out for a certain period, typically from a few days to a few weeks, to be relevant for environmental protection. Sometimes, there is little additional cost involved in monitoring the endpoint over time compared to measuring it only at the end of the experiment. But in spite of the availability of time-resolved data, CECs reported in the literature are mostly based on data at the end of the experiment only. This entails that they are only valid for given exposure times and exposure scenarios. They are not relevant to describe the response to fluctuating contaminant exposures which occur in the field, such as high concentration during the day and low concentration during the night for hospital effluents, or pesticide concentration peaks after a rain event in a river surrounded by treated fields. In this third part of the thesis, we present a concentration-response model with a description of the time-dependence of the reponse. It is based on a Toxico-Kinetic Toxico-Dynamic (TKTD) model published in [Jager et al., 2011]. We apply it to model time-resolved data on the salinity tolerance of riverine species. We use the same hierarchical structure as in the previous part of the thesis and we adapt it for the time-resolved model. We compute a time-resolved SSD usable in principle for all exposure times, and we show how to compute the response of the community for arbitrary contaminant exposure scenarios. This work was developped in collaboration with Ben Kefford (University of Canberra) and Christophe Piscart (Université de Rennes), it was accepted for publication in Environmental Science & Technology.

# 4.1 Time-resolved data

## 4.1.1 Introduction

Toxicity data obtained from bioassay are often time-resolved. Indeed, experiments need to be carried out for a certain period, typically from a few days to a few weeks, to be relevant for environmental protection. When it is possible to easily perform a non-destructive measurement[1] of the target endpoint (e.g. counting the number of survivors), there is little additional cost involved in monitoring the endpoint over time compared to measuring it only at the end of the experiment and indeed the raw bioassay data are often time-resolved. But in spite of the data availability, CECs reported in the literature are mostly based on data at the end of the experiment only. This is due to the fact that concentration-response models rarely describe time-dependence of the response to contaminant exposure. But it entails that end-of-experiment CECs are only valid for given exposure times and exposure scenarios. They are not relevant to describe the response to fluctuating contaminant exposures which occur in the field, such as high concentration during the day and low concentration during the night for hospital effluents, or pesticide concentration peaks after a rain event in a river surrounded by treated fields. Moreover, in classical bioassays exposure times may be short relative to the life cycle of some species. SSD can only hope to be relevant for exposures of that particular time-scale. Another remark concerning the ecological realism of SSD bears on the difficulty to interpret the $HC_p$. Whether an $HC_p$ truly protects $(1-p)\%$ of the community crucially depends on the type of CEC. An $HC_5$ based on $EC_{50}$ is a concentration for which 95% of the species are affected at a level between 0 and 50%. At such an $HC_5$, a community might suffer strong adverse effects if many species are affected at 45%. This encourages the use of CEC for a lower level of effect, such as the $EC_{10}$. However, the estimation of a low level of effect is usually much less precise than that of the $EC_{50}$. In particular, the confidence interval on an $EC_{10}$ obtained from a common log-logistic model is usually larger than on an $EC_{50}$, mostly because the slope at the $EC_{10}$ is not as steep at the $EC_{50}$, as we have seen in the previous part of the thesis. Several authors have advocated that SSDs should be based on NOEC to address this issue. However, the use of this type of CEC has been disparaged extensively because they are based on a wrong interpretation of statistical tests (no statistically significant effect does not mean no effect), they are strongly dependent on the experimental setting and they favour poor resolution on the concentration scale[Warne and Van Dam, 2008, Fox et al., 2012, Chapman et al., 1996, Fox, 2008]. SSD has also been criticized on statistical grounds because the sample of species tested to infer the sensitivity distribution of the community is not representative of known communities and not random[van der Hoeven, 2004], because sample sizes are usually small and because uncertainty on the estimated CEC is generally not taken into account[Kon Kam King et al., 2015, Aldenberg and Rorije, 2013, Moore et al., 2010].
Various approaches have been proposed to improve SSD. Time-dependence has been addressed by extrapolating long term effects from short term exposure through the use of acute-to-chronic transformation[Grist et al., 2006, Mayer et al., 2002,

[1] In the case of the diatom data in the previous part of the thesis, fluorescence was only measured at the end of the experiment.

Duboudin et al., 2004a], or by specifically modelling the toxicity over time with a Dynamic Energy Budget ecotoxicological model (DEBtox)[Smit and Ebbens, 2008, Kooijman and Bedaux, 1996]. The potential implications of taking time-dependence into account have also been investigated, showing that complex patterns can emerge[Fox and Billoir, 2013]. For instance, although species become more sensitive as exposure increases, shifting the mean of the SSD accordingly, they showed that it was also critical to consider the evolution of the variance of the SSD to predict the evolution of the $HC_5$. A solution proposed for the difficulty of interpreting the $HC_5$ was to use NEC models[Fox, 2008, van der Hoeven, 2004, Jager et al., 2011, Kooijman et al., 1996]. Contrary to the $HC_5$ based on a distribution of $EC_x$, the $HC_5$ based on a distribution of NEC can be considered as the threshold below which 95 % of the species suffer no direct effect for the endpoint measured. NECs are estimated as parameters of a threshold model and not from statistical tests, so they do not suffer from the same defects as the NOECs. They represent a threshold below which the concentration has no effect on the endpoint considered.
RTT[Kefford et al., 2005a] has been proposed as a way to overcome some of the statistical issues of SSD, namely small sample size, non randomness and representativity. RTT aims to approximate the sensitivity of a large number of species in order to get a representative sample of the community. The focus of RTT is on the estimation of the variability within the community rather than on the precise estimation of the sensitivity of a few species. Yet, RTT raises the issue of including uncertainty on the tested species sensitivity with increased sharpness. Rapid test data have a large uncertainty which is not included in classical SSD. There have been proposals to include that uncertainty by describing species sensitivity as censored data[Moore et al., 2010, Kon Kam King et al., 2014, Dowse et al., 2013, Kefford et al., 2012a, Hickey et al., 2012], using a frequentist or a bayesian approach[Kon Kam King et al., 2015, Hickey et al., 2012, Craig, 2013]. There is currently no method which addresses all these issues simultaneously. We propose in this part of the thesis a new hierarchical Toxico Dynamic (TD) model for SSD which provides a solution to all the aforementioned problems. It is derived from the TKTD model presented in [Jager et al., 2011], but it aggregates everything related to Toxico Kinetic (TK) in a single compartment (see subsection 4.2.2). We model mortality as a stochastic process resulting from the damage induced by the contaminant to an organism. Contaminant concentration and damage are linked by a phenomenological one-compartment model describing all the biochemical and physiological processes leading to mortality[Jager et al., 2011]. Using this TD model, we model explicitly the time-dependence of individual species sensitivity using parameters which do not depend on time themselves, arriving at a time independent SSD. The TD model is parametrised using a NEC, allowing for a convenient interpretation of the $HC_5$. Additionally, the hierarchical approach allows a consistent management of the statistical uncertainty which is propagated from the original raw data to the $HC_5$. The hierarchical model incorporates all the information available on the tested species regardless of the quality of the data, making it easy to include rare species for which it would not be possible to fit a concentration-response model.

### 4.1.2 Salinity data

The dataset used for this study was published and described in detail[Kefford et al., 2012a]. It contains salinity tolerance of 217 macroinvertebrates taxa from Southern Murray Darling Basin, Victoria[Kefford et al., 2006] and France (Britanny)[Kefford et al., 2012a]. The data consist in the mortality of stream invertebrates exposed to artificial sea water with observations of survival after 24, 48, 72 and sometimes 96h. The salt source for the experiments was artificial sea water: Ocean Nature in Victoria (Aquasonic,Wauchope, NSW, Australia) and Instant Ocean (Red Sea Pharmaceuticals, Haifa, Israel) in France, and experiments confirmed these two brands of artificial sea water had no effect on toxicity[Kefford et al., 2012a]. The only difference in the method used was the temperature at which the experiments were conducted reflecting local climate: 20(± 2)° C in Victoria and 18(± 1)°C in France. All other methods were identical across the studies and are described elsewhere[Kefford et al., 2012a, Kefford et al., 2006]. Multiple species were exposed in the same water but prevented from physically interacting by housing them in enclosures within a larger body of water. All species housed together in this way were collected concurrently from the same site, so if they chemically interacted in the experiments, they likely chemically interacted in the field. Data were collected according to the RTT framework[Kefford et al., 2003] to approximate the sensitivity of large numbers of field-collected taxa. Data coming from Australia and France were pooled together, as the previous study[Kefford et al., 2012a] found that salinity tolerance was more variable across taxonomic group than across region. The dataset is strongly inhomogeneous (Figure 4.1) in the number of concentrations tested per taxa (1 to 98 measured concentration), in the number of organisms tested per concentration (1 to 17) and in the number of replicates per treatment (1 to 3). Notably, there are some common species for which it is possible to estimate an $LC_{50}$ precisely and rare species for which it is only possible to estimate a censored value for the $LC_{50}$ (Figure 4.1). The dataset includes organisms that disappeared either because they were eaten, they completed the aquatic phase of their life-cycle or were otherwise lost and thus could not be followed up (these are hereafter referred to as lost to follow-up organisms).

## 4.2 Toxico-Kinetic Toxico-Dynamic models

### 4.2.1 General intro on Toxico-Kinetic Toxico-Dynamic models

TKTD models in ecotoxicology are compartmental models similar to what can be found in toxicology (Pharmaco-kinetic Pharmaco-dynamic), electric systems, epidemiology (Susceptible Infected Removed models), structured populations in ecology, etc. These models are intended to provide a mechanistic description of contaminant effects, as opposed to empirical models such as the log-logistic concentration-effect model or the various hormesis models presented in the previous section. They account for the multiple processes involved in toxicity, such as entry of the contaminant in the organism, metabolism, dispersion of the metabolites in the organism, build up of damage, lesions and consequently an observable/measurable effect. This effect refers to the concept of

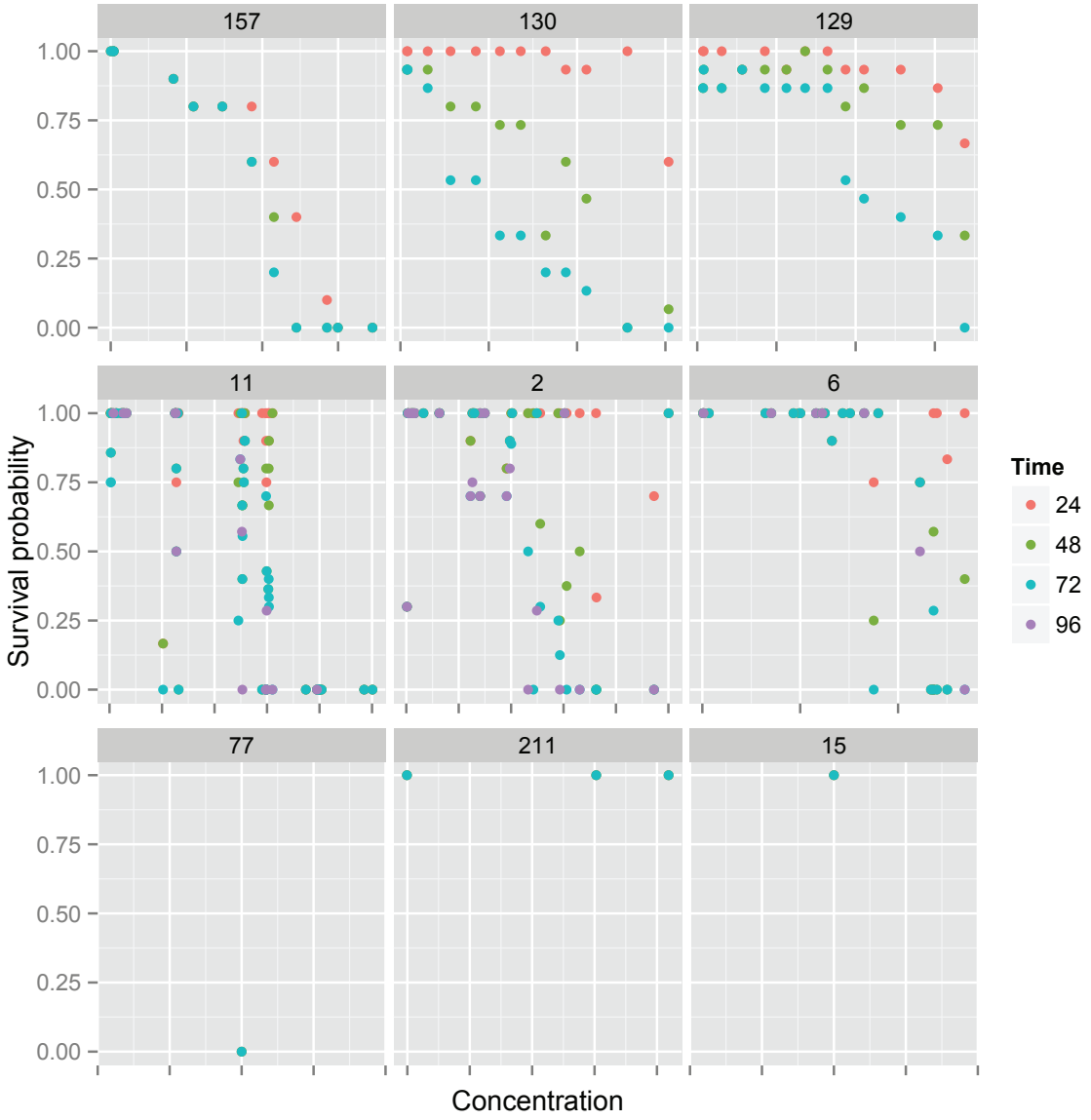



Figure 4.1: 9 out of the 217 species of the salinity dataset. The survival probability is plotted against the salinity for various measurement dates. The top row contains species for which there is an apparent concentration-response relationship and little variability (there is no replicate). The middle row contains species for which there is a lot of data available, but for which variability is very large. The bottom row contains rare species for which very little data is available, so that they do not allow fitting a concentration-response model.

endpoint mentioned in the previous parts of the thesis. Classical endpoints in ecotoxicology include size, body mass, fluorescence, survival, number of offsprings/size of clutch, abnormalities, etc. An important feature of TKTD models is that they include a time-resolved description of these processes. The model parameters are time-invariant, which makes it possible to predict the mortality outside of the time span covered by the data without the uncertainty exploding and to compute time-invariant $HC_5$. In particular, it is possible to compute a Lethal Concentration for $x\%$ of the organisms ($LC_x$) for any time. As noted in [Smit and Ebbens, 2008], in this model any $LC_x$ will decrease asymptotically in time towards the NEC threshold.
[Jager et al., 2011] presented a General Unified Threshold model for Survival (GUTS), a unified framework for the various survival TKTD models in the form of a general model with many parameters from which many TKTD models can be obtained, as limit cases. Their unification is more directed at the Toxico Dynamic (TD) part of the models: they argue that two paradigms prevail for describing the death of an organism of a given species, or more precisely for describing the variability among the response of several organisms of the same species: either death is a deterministic process and there is a variability in the tolerance of the organisms (this is called the Individual Threshold (IT) model, Figure 4.2), or there is no intra-specific variability but death is a stochastic process and variability results from stochasticity (this is called the Stochatic Death (SD) model, Figure 4.2). One way to understand the difference between these two paradigms is to consider the implications of exposing several organisms from the same species to repeated identical concentration pulses of contaminant with a time interval sufficient for complete recovery: in the IT model, only the first pulse will have an effect, wiping out the most sensitive organisms, whereas in the SD model, all pulses will have the same effect on the survival probability but there will be a variability in the number of survivors. The General Unified Threshold model for Survival (GUTS) model incorporates both the IT and SD concepts.

### Toxico-kinetics

We start by presenting briefly the GUTS model to explain why we had to adapt it. In the GUTS model, the TK part can be modelled using a simple one compartment model.

$$\dot{C}^i = k^i C^w - k^e C^i \tag{4.1}$$

where $\dot{C}^i = \frac{dC^i}{dt}$ is the time derivative of the internal concentration, $C^w$ is the external water concentration in contaminant, $C^i$ is the internal concentration, $k^i$ the rate at which the contaminant enters the organism and $k^e$ the rate at which it is eliminated from the organism. In Equation 4.1, $\dot{C}^i$, $C^w$ and $C^i$ depend on time, while the other quantities are time-invariant. If there is information available, this model can be extended to include other TK processes such as bio-transformation or saturation. To integrate Equation 4.1, it is necessary to know the value of $C^i$ at a certain date. The most common assumption is $C^i(t = 0) = 0$. Several exposure scenarios ($C^w$ as a function of time) lead to an analytical formula for the solution $C^i$, notably constant exposure or piecewise constant exposure.

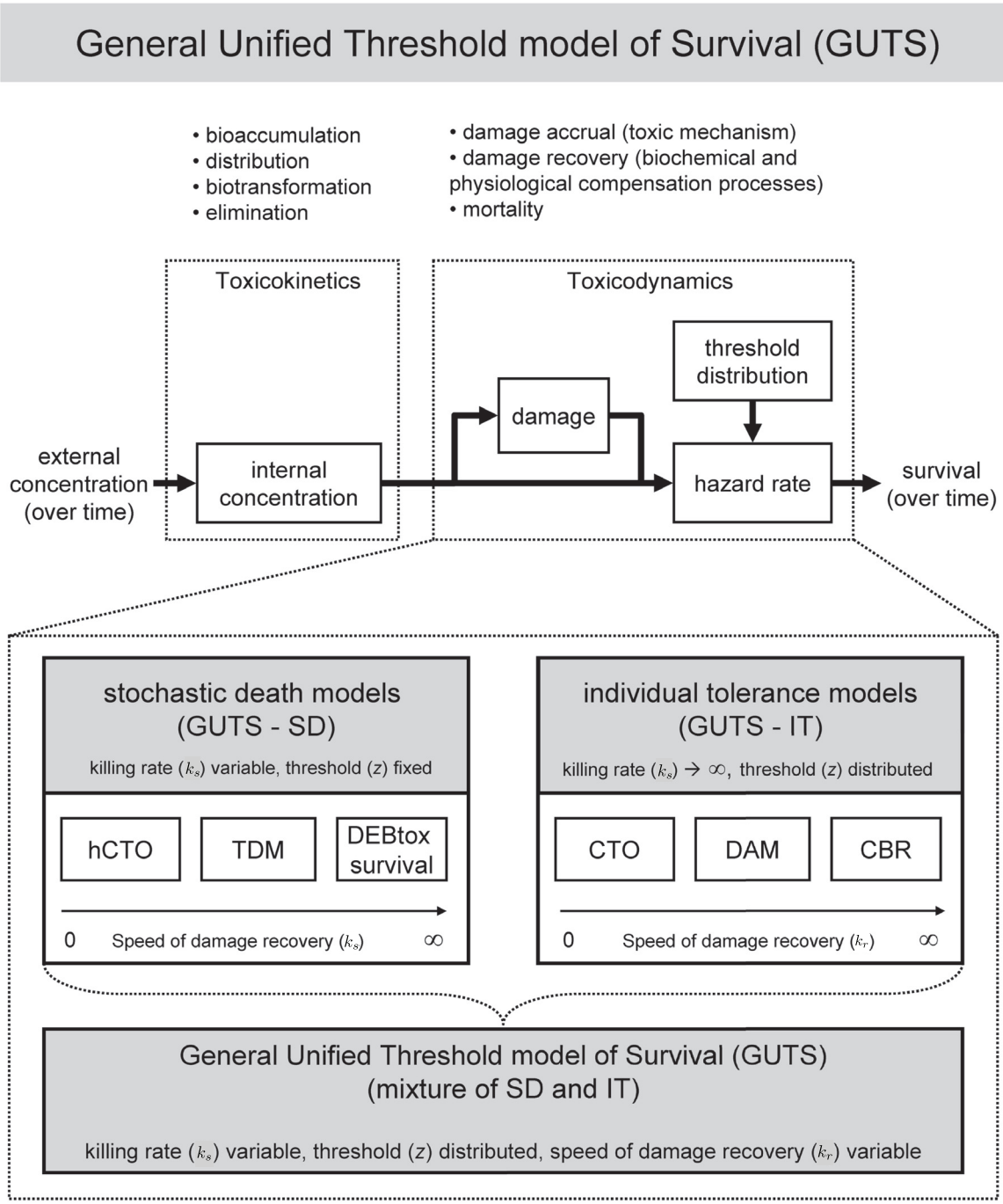



Figure 4.2: Schematic structure of the General Unified Threshold model for Survival (GUTS), a toxicokinetic-toxicodynamic (TKTD) model, consisting of four modules: internal concentration, damage (optional), hazard rate, and threshold distribution. Various existing TKTD models can be derived as special cases within GUTS by making specific toxicodynamic assumptions about the killing rate and the threshold. These special cases consist of stochastic death models (SD) and individual tolerance models (IT), whereas GUTS is a more general, mixed model, comprising both SD and IT. hCTO: hazard (SD) versions of the Critical Target Occupation model; TDM: Threshold Damage Model; DEBtox survival: survival part in the Dynamic Energy Budget models for toxic stress; CTO: Critical Target Occupation model; DAM: Damage Assessment Model; CBR: Critical Body Residue concept. Figure and legend reproduced from [Jager et al., 2011].

### Toxico-dynamics

The TD of the model involve an optional damage compartment to model all the biological, chemical and physiological processes linking internal concentration to toxicity. [Jager et al., 2011] suggest using this damage compartment only when the internal concentration is not sufficient to explain the time-course of mortality.

$$\dot{D} = k^a C^i - k^r D \tag{4.2}$$

where $\dot{D}$ is the time derivative of the damage, $C^i$ is the internal concentration in contaminant, $D$ is the damage, $k^a$ the damage accrual rate and $k^r$ the damage recovery rate. In Equation 4.2, $\dot{D}$, $C^i$ and $D$ depend on time, while the other quantities are time-invariant. To integrate this equation, it is also necessary know the value of $D$ at a certain date. It is possible to make the assumption that the initial damage due to the concentration in contaminant is negligible. It seems reasonable to assume that the organism has not been exposed to hazardous levels of contaminant before the experiment.
The survival model also belongs to the TD part of GUTS. The survival function and the hazard rate are defined by:

$$S(t) = P(T > t) \tag{4.3}$$

$$h_z(t) = -\frac{1}{S(t)}\frac{dS(t)}{dt} \tag{4.4}$$

where $T$ is the date of death of the organism and $S(t)$ is the survival function, or the probability to die after $t$. The hazard rate is the instantaneous probability to die at time $t$ conditional on surviving until then.
In the GUTS model, the hazard rate is supposed to be linearly dependent on the dose-metric above a certain threshold specific to the dose metric and to the organism.

$$h_z = m^0 + k^s_M \left(M - z_M\right)_+ \tag{4.5}$$

$h_z$ is the hazard rate, or the instantaneous probability to die.
The notation $(x)_+$ is a shorthand for $\max(x, 0)$. The $M$ in this model is called the dose-metric, this is the quantity which will control the dynamics of survival. $M$ can be the internal concentration $C^i$, or the damage $D$ if a damage compartment is included. The parameters in this model are $z_M$, an individual threshold below which the dose-metric does not have any impact on the survival of the organism, $k^s_M$, a parameter controlling how the dose metric increases the hazard rate and $m^0$, the control mortality for a dose-metric below the threshold. $z_M$ characterises all the intra-specific variability in the GUTS model, the other parameters are the same for all the organisms of a given species.
The final part of the TD model is the multinomial error model for the number of survivors at each time.
Given an initial number of organisms $N_0$ at the beginning of the experiment, the likelihood of measuring $\boldsymbol{N} = (N_1, \ldots, N_n)$ survivors at time $(t_1, \ldots, t_l)$ given the parameters $\boldsymbol{\theta}$ is given by (derivation in section F.1):

$$f(\boldsymbol{N}|N_0,\boldsymbol{\theta}) = \frac{N_0! S_l^{N_l}}{N_l!} \prod_{k=1}^{l} \frac{(S_{k-1}-S_k)^{N_{k-1}-N_k}}{(N_{k-1}-N_k)!} \tag{4.6}$$

**Limit cases**

The SD limit of the GUTS model is obtained by assuming letting the distribution $z_M$ approach a Dirac distribution, i.e. letting the threshold to be the same for all organisms of a given species. In the SD model, there is no individual variability at the level of the organism and all the variability in the response is attributed to the stochastic variability of the multinomial sampling.

The IT limit of the GUTS model is obtained by letting parameter $k_M^s \to \infty$, which implies that if the dose metric goes above the threshold then the organism dies instantly. Otherwise, toxicity does not have any effect on survival. This removes stochasticity in the death process and all the variability in the response is attributed to individual variability.

### 4.2.2 Adapting the General Unified Threshold model for Survival to model salinity tolerance

We chose a GUTS-SD type model for the salinity dataset. This model is simpler than a full GUTS model because it has less parameters. GUTS-SD is also a little more natural than GUTS-IT, because the IT assumption with deterministic death (GUTS-IT) entails that each organism dies instantly when its threshold is reached, in a sudden step from no effect to effect[Baas et al., 2009]. Instead, the stochastic death assumption entails that the survival probability of an organism decreases approximately exponentially with the concentration over its threshold.

However, we could not use the GUTS-SD model in its exact original form. Our survival data contained no information on the initial internal salt concentration and it could not be assumed to be negligible to solve Equation 4.1, contrary to the xenobiotic contaminants on which the GUTS model was developed. Salts are crucial components of organisms and freshwater species are hyperosmotic, meaning that their internal salinity is greater than that of their environment. As salinity rises they either maintain their internal salinity at the same level (osmoregulate) or osmoconform, meaning that their internal salinity rises with their environment. There are no cases where the initial internal salinity is negligible. This prevented the direct use of the GUTS-SD model described by Jager and colleagues[Jager et al., 2011, Ashauer et al., 2010, Ashauer et al., 2006, Ashauer et al., 2013]. We resorted to modelling damage to the organism instead of internal concentration. This entailed a slightly different understanding of the damage concept than for the GUTS model: damage still describes the state of the organism resulting from all the biochemical and physiological processes which are involved in toxicity[Jager et al., 2011], but it is driven by the external concentration rather than by the internal concentration. More precisely, we used a one-compartment model linking the damage directly to the external concentration, removing the intermediary internal concentration. In that sense, damage aggregates everything related to TK, whereas it was only contained in the TD part of the GUTS model. Hence, it is

not really a TKTD model but a TD model. Entry of the contaminant in the organism is included among all the biochemical and physiological processes lumped into the damage. Nevertheless, this understanding of damage remains fully compatible with the very broad definition given in [Jager et al., 2011]. In our adapted model, mortality rate results from the state of damage of the organism. Appendix G shows that under certain assumptions, the GUTS model and its definition of damage are equivalent to the model presented here, however there are no grounds for making that assumption. It is difficult (albeit possible in principle) to establish what specifically is damage, and it could be specific to the taxa considered. But damage is a concept sufficiently general to allow within a single model to account empirically for a variety of processes.

The compartmental equation for damage is:

$$\dot{D} = k^a C^w - k^r D \tag{4.7}$$

where $D$ is the salinity induced damage, $k^a$ is the damage accrual rate proportional to the external salt concentration $C_w$ and $k^r$ is the damage recovery rate.
Initial damage was assumed to be negligible for all organisms, as all organisms were collected from salinity well below those that induced mortality. Taxa mortality resulting from the capture, the stress from the experiment and everything which does not depend on salinity is accounted for by a specific parameter $m^0$ in the hazard rate.

With these assumptions and for a constant $C^w$ , Equation 4.7 can be integrated as:

$$D = \frac{k^a}{k^r} C^w \left(1 - e^{-k^r t}\right) \tag{4.8}$$

Parameter $k^r$ can be seen as representing the delay between exposure and effect. $\frac{1}{k^r}$ is the characteristic time-scale of the one-compartment model, i.e. the time needed for the damage to reach $e^{-1} = 63\%$ of its maximum value at constant external concentration. In the sense that $D$ is the representation of several TKTD processes modelled with several compartments, $k^r$ is conditioned by the slowest of these processes. This is similar to the notion of rate-determining step in chemical kinetics.

In a standard threshold model, there is a no effect internal concentration $NEC^i$ below which the survival of the organism is not affected by the toxicity (salinity in our case). This translates to a no effect internal damage $NED$ below which the survival of the organism is not affected. Given that $\lim_{t\to\infty} D = \frac{k^a}{k^r} C^w$ it is possible to rescale the $NED$ to units of external concentration and define an external no effect concentration:

$$NEC = \frac{k^r}{k^a} NED \tag{4.9}$$

Next, we define the hazard rate $h_z$ with a linear effect of the damage above the threshold:

$$h_z = m^0 + k_d^s \left(D - NED\right)_+ \tag{4.10}$$
$$= m^0 + k^s \left(C^w \left(1 - e^{-k^r t}\right) - NEC\right)_+ \tag{4.11}$$

where $(x)_+$ means the maximum of $x$ and 0, $m^0$ is the control mortality (for $D \leq NED$) $k_d^s$ is the mortality induced by the damage over the no effect damage threshold and $k^s = \frac{k^a}{k^r} k_d^s$ is called the killing rate, as proposed by Kooijman[Kooijman and Bedaux, 1996]. $k^s$ has the dimension (concentration.time)$^{-1}$ and controls the effect of salinity on survival. Contrasting with Equation 4.5, the hazard rate includes no individual variability, the threshold is common for all organisms of the same species.

There is a time $t_{NEC}$ before which an organism starting with negligible initial internal damage will have its internal damage below the no effect damage threshold. This time can be obtained from Equation 4.8 and Equation 4.9 by :

$$t_{NEC} = -\frac{1}{k^r} \ln\left(1 - \frac{NEC}{C^w}\right) \tag{4.12}$$

Only after that time does the salinity start having an effect on the organism's survival.

The survival probability for one species at time $t > t_{NEC}$ and constant concentration $C^w$ is obtained from the hazard rate by:

$$S(t) = e^{-\int_0^t h(u)du} \tag{4.13}$$
$$= e^{-m^0 t - k^s\left[(C^w - NEC)(t - t_{NEC}) + \frac{C^w}{k^r}\left(e^{-k^r t} - e^{-k^r t_{NEC}}\right)\right]} \tag{4.14}$$

For $t < t_{NEC}$:

$$S(t) = e^{-m^0 t} \tag{4.15}$$

Therefore, the parameters characterising the tolerance of a species are $k^r$, $k^s$, $m^0$ and $NEC$. The assumptions in this model are: 1) there is no intra-specific variability in sensitivity to the contaminant, 2) there is a species-specific damage threshold below which there is no hazard due to the contaminant (this assumption leads to the existence of a species-specific $NEC$, see Jaynes[Jaynes, 2003] for a discussion of this assumption), 3) the hazard rate for a species is proportional to the internal damage above the threshold plus a constant. This means that it has a time-dependent component due to the contaminant plus a constant component due to other causes. This implies that there is no significant hormesis or essentiality. 4) Damage evolves following a one-compartment model, it increases proportionally to the external concentration and decreases proportionally to the damage level. 5) The mechanistic TKTD processes that cause acute and chronic toxicity are the same and the greater sensitivity often observed with expanded exposure is the result of a build-up of damage.

### 4.2.3 More details about the interpretation of the Toxico-Dynamic model parameters

The TKTD model describes a monotonous decrease of the survival probability both with time and concentration. The survival probability can be represented as a function of time and as a function of concentration (Figure 4.3). This illustrates the effect of parameter $k^r$, as the models with and without $k^r$ show different behaviours.
In the model without $k^r$, ie. $k^r \to \infty \implies t_{NEC} \to 0$, there is no delay between exposure and effect. The consequences on the time representation (Figure 4.3 a)) is that the survival probability is exponentially decreasing with a characteristic time dependent on the concentration. The consequences on the concentration representation are that the survival probability is flat for concentrations below the $NEC$ and exponentially decreasing for concentrations above the NEC (Figure 4.3 c)).
For a model with $k^r$, there is a delay between exposure and effect. When the external concentration is below the $NEC$, only parameter $m^0$ has an influence on survival for the lowest concentration. When the external concentration is above the NEC, there is a delay $t_{NEC}$ before the internal damage reaches a level affecting the survival of the organism. This delay depends on parameter $k^r$, NEC and on the external concentration (Equation 4.12). After this delay, the effect of salinity increases with time and saturates. On the time representation, this results in a change of concavity (Figure 4.3, b) ). On the concentration representation, this results in a flat region of decreasing length which is above the $NEC$ Figure 4.3 d). The flat part of the curve gets shorter and shorter with time: this is because for any given time, there is an external concentration $NEC_t$ below which the internal damage cannot reach the $NED$ before $t$, and this $NEC_t$ decreases. $NEC_t$ is the external concentration such that starting at zero damage, the damage can exactly reach the $NED$. For external concentrations below $NEC_t$, there cannot be any effect of the concentration on survival at time $t$. We can find an analytical expression for $NEC_t$ starting from Equation 4.8 and using Equation 4.9:
for $C^w = NEC_t$,

$$D(t) = NED \tag{4.16}$$

$$\implies NED = \frac{k^a}{k^r} NEC_t (1 - e^{-k^r t}) \tag{4.17}$$

$$\implies NEC_t = \frac{k^r}{k^a} \frac{NED}{1 - e^{-k^r t}} \tag{4.18}$$

$$= \frac{NEC}{1 - e^{-k^r t}} \tag{4.19}$$

$NEC_t$ decreases with time as a logistic curve and $\lim_{t \to \infty} NEC_t = NEC$.

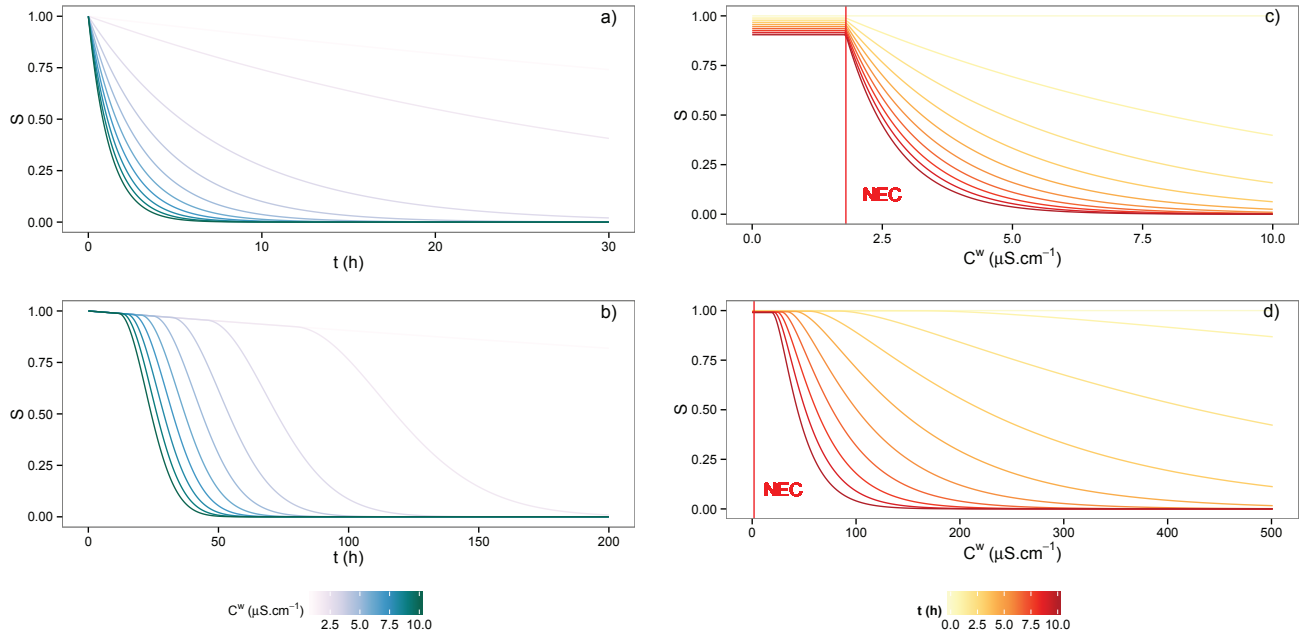



Figure 4.3: a) and b): survival probability as a function of time for various concentrations and c), d): as a function of the concentration for various times. Top panel shows a model without $k^r$, ie. $k^r \to \infty \implies t_{NEC} \to 0$. Bottom panel shows a model with a small value of $k^r$. The scales and parameter values were chosen in order to show the phenomenology of the model but do not correspond to biologically realistic or meaningful values.

### 4.2.4 Error model for survival data

[Jager et al., 2011] used a multinomial formulation of the error model on the number of survivors. The multinomial and the conditional binomial error models[Forfait-Dubuc et al., 2012] are two mathematically equivalent formulations of the same error model for survival without lost to follow-up organisms (see Appendix F). The presence of lost to follow-up organisms precludes the use of the multinomial model formulation and suggests instead modelling the number of survivors at time $t_k$ conditionally to the number of organisms alive at time $t_{k-1}$ and which have not disappeared between $t_k$ and $t_{k-1}$. Replacing the multinomial error model in the presence of lost to follow-up organisms makes the assumption that the mechanism by which they are lost is not related to toxicity (does not influence the estimation of the parameters). The probability for species $j$ exposed to concentration $C_j^w$ to survive until $t_k$ after having lived until $t_{k-1}$ is:

$$S_{i,j}\left(t_k | t_{k-1}\right) = \frac{S_{i,j}\left(t_k\right)}{S_{i,j}\left(t_{k-1}\right)} \tag{4.20}$$

where the dependence on $C_j^w$ was omitted to simplify the notation. We modelled the observed number of survivors $N_{i,j,k}$ with the following binomial distribution:

$$N_{i,j,k} \sim \mathcal{B}\left(N^{p}_{i,j,k}, S_{i,j}(t_k|t_{k-1})\right) \tag{4.21}$$

where $N^{p}_{i,j,k}$ is the number of organisms alive at the previous time and which have not disappeared since. Note that $N^{p}_{i,j,k} \neq N_{i,j,k-1}$ because some organisms might have disappeared between $t_{k-1}$ and $t_k$.

## 4.3 Hierarchical Toxico-Kinetic Toxico-Dynamic model

### 4.3.1 Description of the model

The structure of the hierarchical model was built in the same spirit as that described in subsection 3.4.1 for the diatom data. There, we modelled the multivariate distribution of two parameters of the concentration-effect model. In this part of the thesis we modelled the multivariate distribution of the four parameters of the concentration-response model. The main difference lied in the modelling of the variance-covariance matrix and in the interpretation of the parameters. The aim of this hierarchical modelling still was to include all the information available in the raw data and specifically the uncertainty from the fit of the model. In the particular case of RTT of rare species, this uncertainty can be very large and ignoring it is not ideal. The shrinkage property (pooling of information across species which shrinks together the model parameter estimates) of the hierarchical SSD model was a convenient way of dealing with the very heterogeneous salinity data (Figure 4.1), allowing each species to impact the estimation of the community parameters at the extent of the information available on that species. The two levels of hierarchy were again the species and the community (Figure 4.4 and Table 4.1), where the species play the same role as random effects in a mixed model. As the log-normal distribution is the distribution most frequently used in classical SSD[Posthuma et al., 2010], we followed that custom and assumed a log-normal distribution for each of the parameters. This is the same assumption as made for the hierarchical modelling of the diatom data (chapter 3). We also modelled a potential correlation among the parameters with multivariate normal correlation matrix.
$\mu$ and $\Sigma$ are defined in terms of position, scale and correlation parameters for each of the four dimensions of the multivariate normal distribution:

$$\mu = (\mu_{\log k^r}, \mu_{\log k^s}, \mu_{\log m^0}, \mu_{\log NEC}) \tag{4.22}$$

$$\Sigma = \xi \Lambda \xi \tag{4.23}$$

where:

$$\xi = diag(\sigma_{\log k^r}, \sigma_{\log k^s}, \sigma_{\log m^0}, \sigma_{\log NEC}) \tag{4.24}$$

and

$$\Lambda = \begin{pmatrix} 1 & \rho_{\log k^r, \log k^s} & \rho_{\log k^r, \log m^0} & \rho_{\log k^r, \log NEC} \\ \rho_{\log k^r, \log k^s} & 1 & \rho_{\log k^s, \log m^0} & \rho_{\log k^s, \log NEC} \\ \rho_{\log k^r, \log m^0} & \rho_{\log k^s, \log m^0} & 1 & \rho_{\log m^0, \log NEC} \\ \rho_{\log k^r, \log NEC} & \rho_{\log k^s, \log NEC} & \rho_{\log m^0, \log NEC} & 1 \end{pmatrix} \tag{4.25}$$

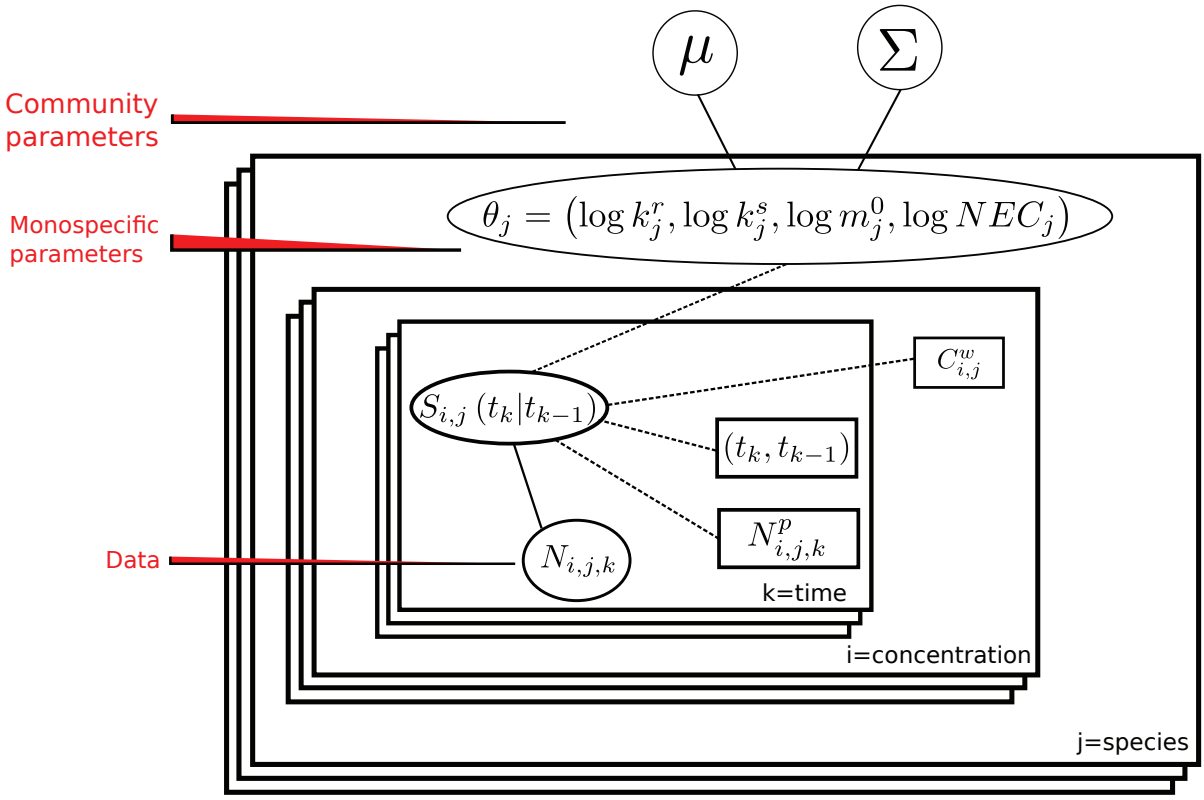



Figure 4.4: Probabilistic directed acyclic graphical model of the hierarchical model[Koller and Friedman, 2009] (also called Bayesian network). Ellipses represent variables, rectangles represent covariables. Solid lines represent stochastic links, dotted lines represent deterministic links. To avoid repetition, similar subunit are summarized by plates. The inner set of plates denote the different times, the intermediate set of plates the different concentrations and the outer set of plates denote the different species. Since the graph is directed and acyclic, any two variables are conditionally independent given the value of their parents. Links are detailed in Table 4.1.

### 4.3.2 Modelling the variance-covariance matrix

One of the main difference between this model and that presented in subsection 3.4.1 lies in the modelling of the variance-covariance matrix. It is not possible to use the same method to model a 2×2 variance covariance matrix and a 4×4 variance covariance matrix.

Table 4.1: Description of the links in probabilistic directed acyclic graphical model (Figure 4.4). $S_{i,j}$ is defined from Equation 4.20 and Equation 4.12. $\mu$ and $\Sigma$ are the position and scale parameters of the 4-dimensional multivariate normal distribution, defined in Equation 4.22 and Equation 4.23.

| Node | Type | Equation |
|---|---|---|
| $\theta_j$ | Stochastic | $\theta_j \sim \mathcal{N}_m(\mu, \Sigma)$ |
| $S_{i,j}\left(t_k \middle| t_{k-1}\right)$ | Deterministic | $S_{i,j}\left(t_k \middle| t_{k-1}\right) = \frac{S_{i,j}(t_k)}{S_{i,j}(t_{k-1})}$ |
| $N_{i,j,k}$ | Stochastic | $N_{i,j,k} \sim \mathcal{B}\left(N^p_{i,j,k}, S_{i,j}(t_k \middle| t_{k-1})\right)$ |

This is because a variance covariance matrix is constrained to be positive-definite[2]. For the $2 \times 2$ variance covariance matrix, positive scale and correlation parameter between 0 and 1 automatically produce a positive-definite matrix. For a $3 \times 3$ matrix, it is possible to generate two correlation parameters at random and to derive a interval where the last correlation parameter must lie to satisfy positive-definitiveness. For higher dimensions, it is not practical to proceed by correlation parameters, the most common method is to sample correlation matrices from an Inverse Wishart distribution. However, this method has been criticized because it introduces a correlation between the scale and the correlation parameters in the prior, which is generally not a desirable feature for modelling. Moreover, to determine a prior on a variance covariance matrix, it is more natural to give separate priors on the scale and correlation parameters because they have a clearer interpretation. Barnard, McCulloch and Meng[Barnard et al., 2000] proposed to use a separation strategy and divide the variance-covariance matrix in a correlation matrix and a scale matrix (Equation 4.23). Using this formulation, it is possible to specify priors separately on each scale parameter and on the correlation matrix. One type of prior using this strategy is the scaled Inverse Wishart prior[3][Gelman and Hill, 2007], however there remains some correlation in the prior (Figure 4.5). [Lewandowski et al., 2009] proposed an efficient method to sample from the space of positive definite correlation matrices, which allows to use a uniform prior on the space of all correlation matrices. Using this last method, there is no undesirable correlation among the scale and correlation parameters in the prior (illustration on a $2 \times 2$ matrix on Figure 4.5). The Lewandowski-Kurowicka-Joe [Lewandowski et al., 2009] (LKJ) prior has a $\nu$ parameter which allows to make the prior informative: for $\nu = 1$, the prior is uniform on the space of correlation matrices, for $\nu > 1$ the prior favours small marginal correlations parameters while for $\nu < 1$ the prior favours large marginal correlation parameters. We chose to use of a uniform prior on the space of all correlation matrices was made because we did not expect a priori any particular correlation among the parameters. In a four dimensional space, this uniform prior tends to favour small marginal correlation parameters against large ones, because there are more positive-definite $4 \times 4$ matrices with small marginal

[2] A symmetric $n \times n$ real matrix $M$ is said to be positive-definite if $\forall x \in \mathbb{N}^n \setminus \{0\}^n ; x^T M x > 0$. Another characterisation is that all the eigen values (which are all real) are stricly positive.

[3] The separation strategy for the scaled Inverse Wishart prior is to decompose the variance-covariance matrix into a scale matrix and an unscaled variance covariance matrix rather than into a scale matrix and a correlation matrix.

correlation parameters[Barnard et al., 2000] (see also Figure 4.6).

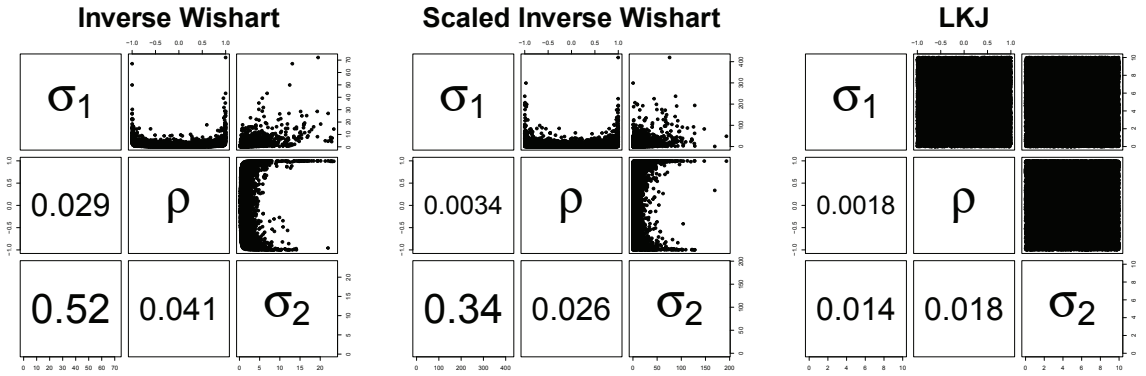



Figure 4.5: Comparaison between the Inverse Wishart and the LKJ prior for the correlation matrices. Scatter plot and Pearson correlation coefficient between the parameters of a $2 \times 2$ variance-covariance matrix generated using the two priors. The idea for this figure was inspired by `https://dahtah.wordpress.com/2012/03/07/why-an-inverse-wishart-prior-may-not-be-such-a-good-idea/` (last accessed on July 14$^{\text{th}}$ 2015).

### 4.3.3 Fit of the model

The model was implemented using the software Stan[Gelman, 2014] via the R interface RStan[Gelman, 2014]. As JAGS[Plummer, 2013] or WinBUGS[Lunn et al., 2000], Stan performs Bayesian inference but it uses hamiltonian Monte Carlo with the No-U-turn sampler[Hoffman and Gelman, 2011] instead of Markov Chain Monte Carlo Gibbs sampling. Stan was chosen for convenience in writing the threshold model and for the availability of various methods to define the prior on the correlation matrix. Hamiltonian Monte Carlo is also expected to perform better than Markov Chain Monte Carlo in terms of autocorrelation in the chains. We used four chains with 5000 warm-up iterations and 5000 post warm-up draws per chain. We chose vague uniform priors for the scale parameters of the normal distributions: $\sigma_{\log k^r}, \sigma_{\log k^s}, \sigma_{\log m^0}, \sigma_{\log NEC} \sim \mathcal{U}(0, 5)$. This strategy is recommended by Gelman[Gelman, 2006] in cases where a lot of data is available for estimating the scale parameters. We also chose vague uniform priors for the position parameters: $\mu_{\log k^r}, \mu_{\log k^s}, \mu_{\log m^0} \sim \mathcal{U}(-7, 2)$. The prior on the $\mu_{\log NEC}$ was chosen to be a little larger than the concentration range, because for many species the concentration range did no contain the NEC (no observable increase of mortality for the highest tested concentration or 100% mortality at the lowest concentration, see also Figure 4.1):

$$\mu_{\log NEC} \sim \mathcal{U}\left(\log \min_{i,j}\left(C_{i,j}\right) - 1, \log \max_{i,j}\left(C_{i,j}\right) + 1\right) \tag{4.26}$$

Note that the log-normal marginal distribution of the NEC for each species implies that the NEC can be arbitrarily small. The Stan script for fitting the model is provided in Appendix H.

### 4.3.4 Results of the fit

#### Model convergence

Model convergence was monitored using the split potential-scale-reduction statistic $\hat{R}$[Gelman and Rubin, 1992], ensuring that $\hat{R} \leq 1.1$ for all the nodes as recommended by Brooks and Gelman[Brooks and Gelman, 1998]. Comparison between priors and posteriors showed a clear shrinkage, sign that there was sufficient information in the data to estimate the parameters (Figure 4.6).

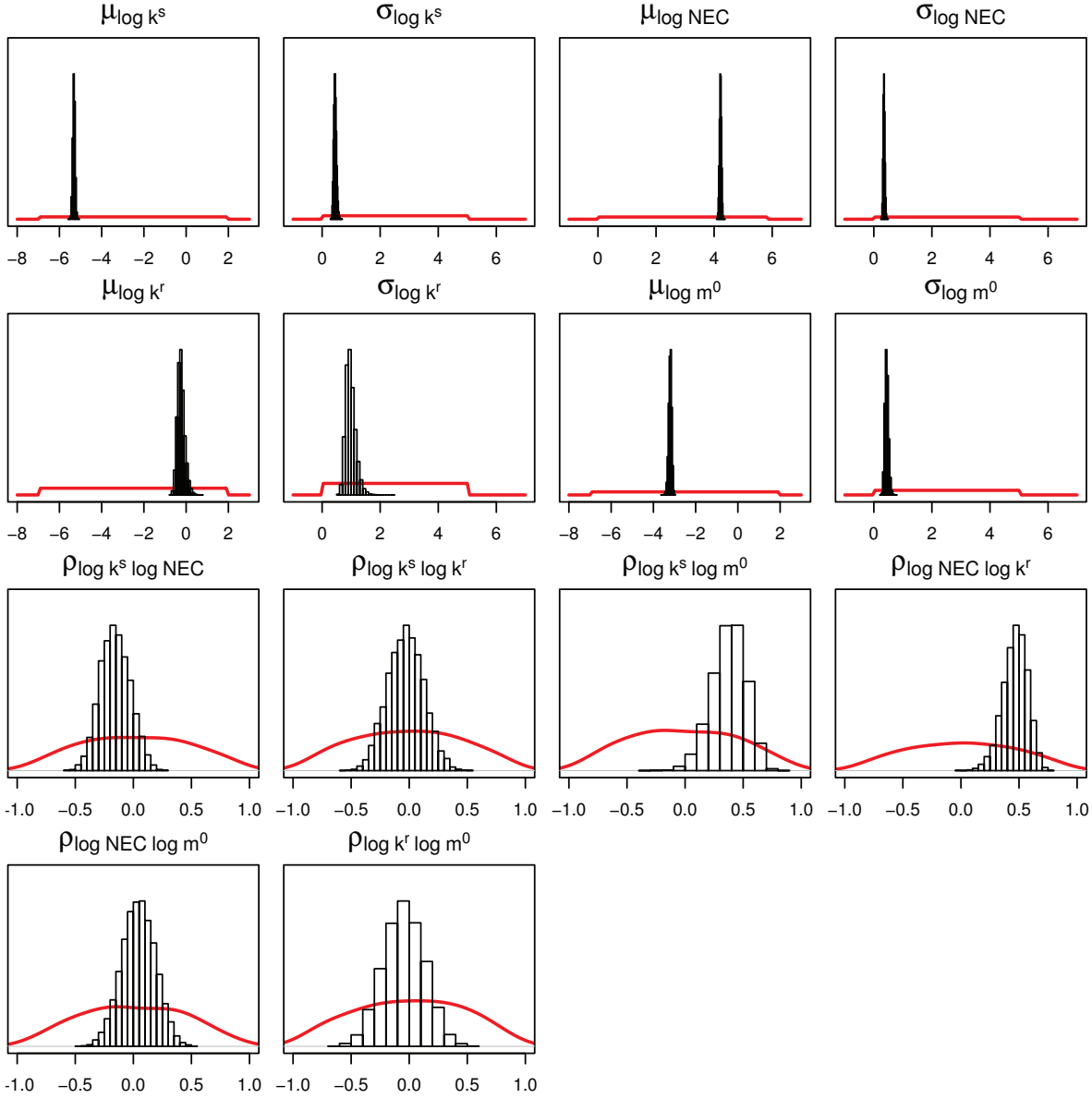



Figure 4.6: Comparison between prior distributions (red) and posterior distributions (black histogram). The marginal priors on the correlation parameters are obtained from the Lewandowski-Kurowicka-Joe distribution. The prior is uniform over the space of the $4 \times 4$ correlation matrices.

**Estimated parameters**

In a classical SSD, only variability on the $EC_x$ or the NEC (via parameter $\sigma_{\log NEC}$) would be considered. All other parameters would be treated as if they were constant. However, our hierarchical model shows that the variability on the other parameters is similar if not greater than the variability on the NEC (Table 4.2). Particularly, there seems to be a large variability on $k^r$, the parameter controlling the delay between exposure and effect. We found many species with a large value of $k^r$, which corresponds to a rapid equilibration between the internal damage and the water concentration. It could be tempting to think that for many species, there is no delay between exposure and effect. However, the survival function depends on the whole time course of the internal damage (Equation 4.13) and for large external concentrations, a lot can happen even in the small period it takes for the concentrations to equilibrate. We also observed a mild positive correlation between parameters $k^r$ and NEC and between parameters $k^s$ and $m^0$. It means that the tolerant species tend to have a short delay between effect and exposure, and that the species with large control mortality (who might have suffered a strong stress from the capture and experimental setting) tend to suffer a strong effect of the concentration.

Table 4.2: Estimated community parameters and their 95% credible interval. The estimate is the median of the posterior distribution.

| Position | Estimate | Scale | Estimate | Correlation | Estimate |
|---|---|---|---|---|---|
| $\mu_{\log k^r}$ | -0.26 [-0.63,0.33] | $\sigma_{\log k^r}$ | 1.0 [0.72,1.6] | $\rho_{\log k^r \log k^s}$ | 0.12 [-0.43,0.23] |
| $\mu_{\log k^s}$ | -5.3 [-5.4,-5.2] | $\sigma_{\log k^s}$ | 0.39 [0.31,0.50] | $\rho_{\log k^r \log m^0}$ | -0.15 [-0.50,0.23] |
| $\mu_{\log m^0}$ | -3.1 [-3.3,-3.0] | $\sigma_{\log m^0}$ | 0.42 [0.30,0.58] | $\rho_{\log k^r \log NEC}$ | 0.53 [0.25,0.73] |
| $\mu_{\log NEC}$ | 4.1 [4.0,4.3] | $\sigma_{\log NEC}$ | 0.38 [0.30,0.49] | $\rho_{\log k^s \log m^0}$ | 0.29 [0.05,0.57] |
| | | | | $\rho_{\log k^s \log NEC}$ | -0.20 [-0.48, 0.10] |
| | | | | $\rho_{\log m^0 \log NEC}$ | 0.01 [-0.32,0.32] |

**Robustness to a change of vaguely informative priors**

Gelman reported that the estimate for the scale parameter in a hierarchical model with a uniform prior can be sensitive to the higher bound of the support[Gelman, 2006]. To check if this was the case, we tested $\sigma_{\log k^r}, \sigma_{\log k^s}, \sigma_{\log m^0}, \sigma_{\log NEC} \sim \mathcal{U}(0,4)$ and $\sigma_{\log k^r}, \sigma_{\log k^s}, \sigma_{\log m^0}, \sigma_{\log NEC} \sim \mathcal{U}(0,6)$. We found no appreciable difference in the posterior distributions. Similarly, we studied the impact of changing the prior on the correlation matrices to favour large correlations ($\nu = 2, \nu = 4$) or small correlations ($\nu = 0.25, \nu = 0.5$) and found no appreciable difference in the posterior distributions.

### 4.3.5 Assessment of model fit

The fit of the model was assessed by comparing the agreement between the predicted and the observed number of survivors at each time and concentration. The predicted number of survivors was obtained in a typical posterior-predictive check approach[Gelman et al., 2014], by drawing 2000 sets of model parameters from the posterior distribution and simulating 500 draws from the binomial error model for each of

the parameter sets, using $C^w$ and $N^p$ as covariates. Most observations in the dataset seemed to sit well along the identity line (Figure 4.7).

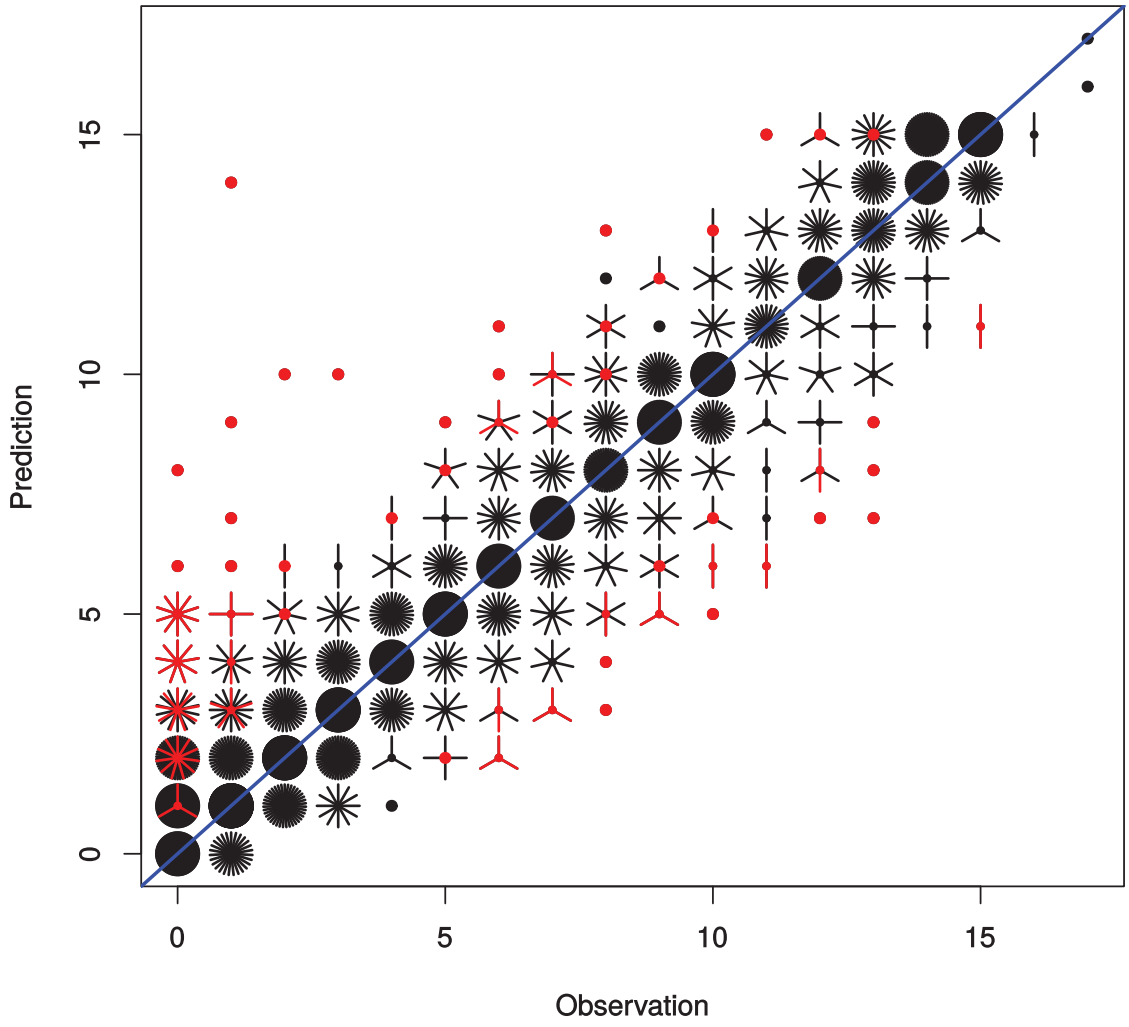



Figure 4.7: Sunflower plot of the predicted number of survivors against the observed number of survivors. A dot on the figure denotes the first data point and each petal of a sunflower denotes an additional data point at the same location. The black colour denotes points for which the 95% scredible interval around the prediction (i.e. the interval between the 2.5$^{\text{th}}$ and the 97.5$^{\text{th}}$ percentiles of the posterior distribution of the number of survivors) contains the observed data, the red colour those for which this is not the case. The blue line denotes the identity line. Therefore, points for which the prediction is equal to the observation lie on the line.

The proportion of points for which the prediction differs from the observation by more

than 6 is only 0.2%. To probe further, we tried to look at the points that were not predicted very well and appeared as outliers on Figure 4.7. Some of these points are on the top left corner of Figure 4.7. To select outliers more objectively, we selected the points for which prediction and data were at 6 units apart and looked at them. Figure 4.8 shows the points (in black) which appear to be outliers in the sunflower plot among the rest of the raw data for the same time and species. These outliers clearly appear to be far from the bulk of the points, showing excessive mortality in all cases but one. Since we did not consider intra-species variability/inter experiment variability in our model, the prediction was expected to fail in these cases. The only variability that our model allows for a given concentration, species and time is the variability from the binomial error model. However, it can be noted that there are only 8 out of 4764 (0.2%) points for which the prediction and the data differ by 6 or more, which suggests that the model is performing well overall. There are 26 points which differ by at least 5 (0.6%), while there are 4 points which differ by at least 7 (0.1%).
The coverage ratio of the 95% credible intervals is the proportion of the observation which are included in the 95% credible intervals around the prediction. If the error model is correct, the expected proportion of the data included in the 95% credible intervals is 95%. Strong deviations from this ratio indicate that the error model is false. For our model, the 95% coverage ratio of the credible intervals was of 97 %, close enough to 95% to deem the error model reasonable.

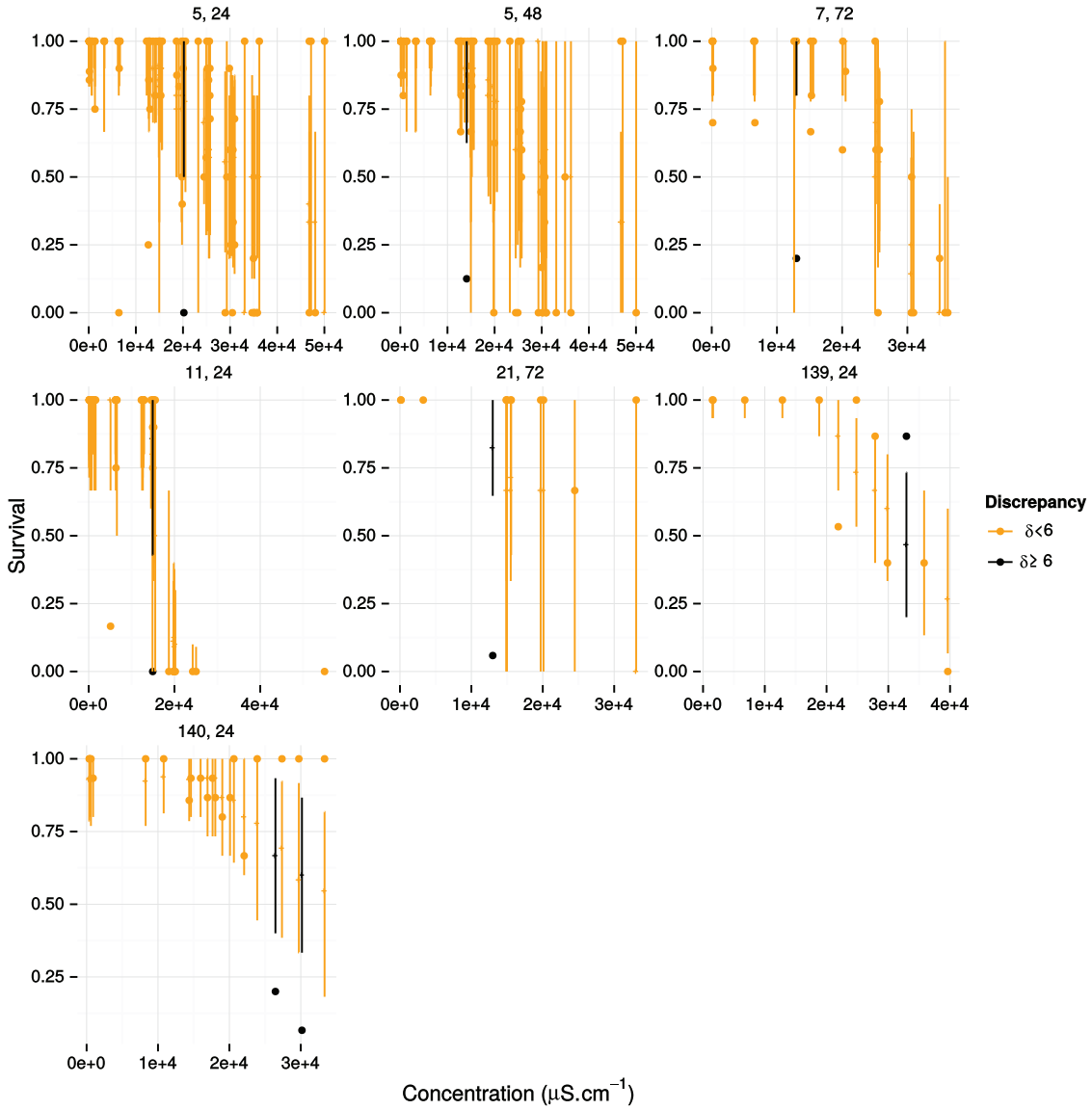



Figure 4.8: Raw data (points) and median prediction of the model (horizontal dash) with the credible interval (vertical bars). This figure place the outliers of Figure 4.7 in the context of the raw data available for the same species and time. The points for which the discrepancy between the prediction and the data is superior or equal to 6 appear in black. The other points appear in yellow. The credible intervals around the prediction were obtained by simulation from the posterior distributions and follow the same colour code as the points.

## 4.4 Comparison with classical Species Sensitivity Distribution

### 4.4.1 Methodology for the comparison

In order to show the added value of our hierarchical approach, we compared it to the classical SSD. To fit the classical SSD, we followed the previous approach by Kefford et al.[Kefford et al., 2003] and fitted a two-parameters log-logistic model to the survival data at 24, 48 and 72 hours using only widely available and user-friendly tools to obtain $LC_{50}$ on which we could compute an $HC_5$ at 24, 48 and 72 hours. The data at 96h were not used for the classical SSD because they concerned much fewer species. For species with sufficient data, we were able to estimate the $LC_{50}$ and a 95% confidence interval with the delta method[Casella and Berger, 2002] using the R package *drc*[Ritz and Streibig, 2005]. For species with insufficient data, or when the delta method gave incoherent confidence intervals, a range for the $LC_{50}$ was estimated by expert judgement from the raw data. For a fairer comparison between classical and hierarchical SSD, some of the model fit uncertainty was taken into account in classical SSD by estimating the $LC_{50}$ as interval-censored data, the interval being defined by the 95% confidence interval. A log-normal SSD was fitted by frequentist maximum likelihood on all the censored $LC_{50}$ using the web-interface MOSAIC_SSD[Kon Kam King et al., 2014] (`http://pbil.univ-lyon1.fr/software/mosaic/ssd/`). The interface computes the $HC_5$ and estimates its confidence interval by bootstrap. Classical SSD is based on $LC_x$ which depend on time. Therefore, it is itself time-dependent and can be computed at 24, 48 and 72 hours. The time-resolved SSD can be based on $LC_x$ obtained from the hierarchical model at 24, 48 and 72 hours as well, but it can also be based on the NEC obtained from the hierarchical model. Since the NEC is time-independent, the SSD based on NEC is also time-independent. In the hierarchical TD model, it is possible to compute numerically the $LC_{50}$ of a species for each time. On these computed $LC_{50}$, it is possible to fit an SSD and to estimate an $HC_5$ for each time. We computed the $HC_5$ as a function of time by sampling 1000 community parameters from the joint posterior distribution, then by simulating 5000 $LC_{50}$ for each of them and calculating the fifth percentile. The 5000 $LC_{50}$ were used for computing accurately the $HC_5$, while the 1000 parameters were used for computing the uncertainty around the $HC_5$. We also computed a time-independent $HC_5$ based on the NEC and its credible interval from the joint posterior distribution of $\mu_{\log NEC}$ and $\sigma_{\log NEC}$.

### 4.4.2 Results of the comparison

The $HC_5$ of the classical SSD and the time-resolved $HC_5$ both computed on $LC_{50}$ decline with increasing exposure time, while the $HC_5$ based on the NEC is constant through time (Figure 4.9). The $HC_5$ based on the NEC ($4407[3310 - 5634]\mu S/cm$) is smaller than the classical $HC_5$ at 24, 48 and 72 H ($10242[8106 - 12805]$, $8821[7112 - 10846]$ and $6800[5405 - 8643]\mu S/cm$, respectively). Deciding to make management decisions based on $HC_5$ calculated from the $LC_{50}$ calculated at the end of the experiment (72 hours) or calculated from the NEC would be of practical importance. The time-resolved $HC_5$ com-

puted on $LC_{50}$ are very similar to the classical $HC_5$: the hierarchical NEC model seems to reproduce the results of the classical approach for $HC_5$ computed at 24, 48 and 72h. The credible intervals around the $HC_5$ are slightly smaller than the confidence intervals on the classical $HC_5$ which reveals that taking into account all the information in the raw data including the uncertainty on the $LC_{50}$ estimate may actually increase the precision on the $HC_5$. Finally, as any $LC_x$ converges towards the NEC[Smit and Ebbens, 2008], the $HC_5$ based on $LC_{50}$ also converges to the $HC_5$ based on the NEC (Figure 4.9).

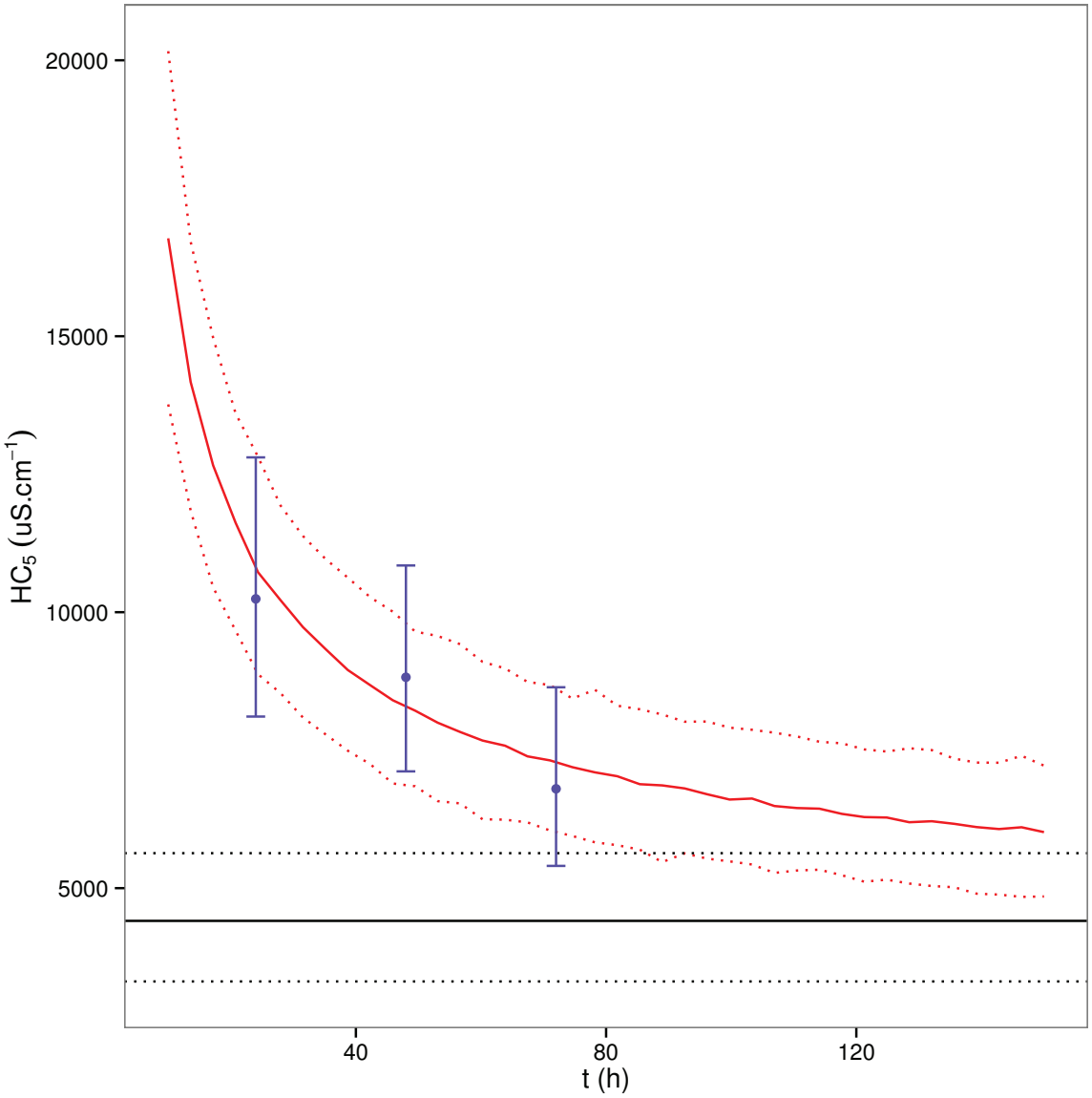



Figure 4.9: Classical $HC_5$ (blue) and hierarchical $HC_5$ (red) computed on the $LC_{50}$ as a function of time and hierarchical $HC_5$ computed on the NEC (black). Vertical segments delimit the 95% confidence interval on the classical $HC_5$. Dotted lines delimit the 95% credible intervals on the hierarchical $HC_5$, solid lines represent the median of the posterior distribution.

## 4.5 Discussion

### 4.5.1 Principal findings

We were able to construct a time-resolved SSD by explicitly modelling the time component of contaminant-induced mortality using a TKTD model. Our hierarchical TKTD approach showed that although classical SSD only considers variability on the CEC, there is a large variability on the other parameters of the concentration-response model in a community and potential correlations. We showed that the $HC_5$ calculated on $LC_{50}$ was time-dependent and that 72h exposure experiments were unable to predict longer term toxicity. We showed how to obtain a time-independent $HC_5$ based on NEC whose implications for the fraction of species affected are easier to interpret, because at the $HC_5$, salinity should have no effect on 95% of the species and this does not depend on the exposure duration in the original experiment. Furthermore, we included data from RTT which constituted a representative sample of the sensitivity from real communities, while accounting properly for the large uncertainty associated with this testing method.

### 4.5.2 Comment on model generalisation

We presented a TD model with only two compartments (i.e. internal damage to the organism and the surrounding water), but the framework we used is very generic. Jager et al.[Jager et al., 2011] contended that more complex compartmental models, if the compartments are not observed, could be rescaled to the two-compartment model because the slowest compensating process (which can be a TK or TD recovery process) would dominate the dynamics of toxicity. Only the interpretation of parameter $k^s$ would change. Here, we opted for a very generic concentration-response model, so as to cover the whole diversity of biochemical and physiological phenomena leading to salinity-induced mortality. It is a requirement of the hierarchical approach that all species should be described with the same concentration-response model, or at least by nested models. If additional data were available to describe some underlying TK or TD processes, it would be possible to move towards a more complex concentration-response model while keeping the hierarchical structure.

### 4.5.3 Methodological implications

The comparison between the classical $HC_5$ calculated at 72h and the time-resolved $HC_5$ calculated using the whole time course of the experiment shows the added value of including all the available data in the SSD analysis: there is no need to use a snapshot for an arbitrary exposure period when there is sufficient data to estimate a time-independent $HC_5$. As discussed in [Kon Kam King et al., 2015], the hierarchical structure of the model entails a shrinkage behaviour at the species level. For instance, the assumption of log-normality of the NEC distribution has a two-fold bearing on the NEC estimate for each species: 1) it is possible to estimate a NEC and its credible interval for rare species and 2) the NEC for the most sensitive/tolerant species will be shrunk in the direction of the intermediate species. Consequently, these species

will have a NEC slightly different from what they would have had without this assumption. Our approach renders more explicit the parametric assumption behind SSD. Our time-resolved SSD model using RTT data also makes the traditional SSD assumption that all tested species are randomly sampled. This is equivalent to making the assumption of species exchangeability([Craig, 2013, Craig et al., 2012] and Appendix B.2), which means that there is no taxonomic or trait-based structure to sensitivity. This assumption has been questioned and proposals have been made to model non-exchangeability[Craig, 2013, Craig et al., 2012], but they would require raw data from other contaminants for the same species, which are not yet available.
Modelling the distribution of all the parameters in the concentration-response model entails more distributional choices than in classical SSD. We defaulted to the same log-normal distribution for all parameters. The estimated random effects showed that it was a generally reasonable choice, although there were some slight departures from log-normality (Figure 4.10). Notably, there was a hint for the presence of a mixture of distributions for parameter $m^0$ and $k^r$, with a separation between those greater and smaller than 0.3 for parameter $k^r$. This corresponds to an equilibration time between internal damage and external concentration of the order of 3 hours. The reason for this mixture of distributions remains to be explored. Future work could develop the model and divide the species into different sub-groups to take this into account. This illustrates the fact that using a mechanistic description of toxic effects of salinity rather than an empirical concentration-response model requires more modelling choices but offers potentially richer interpretations.

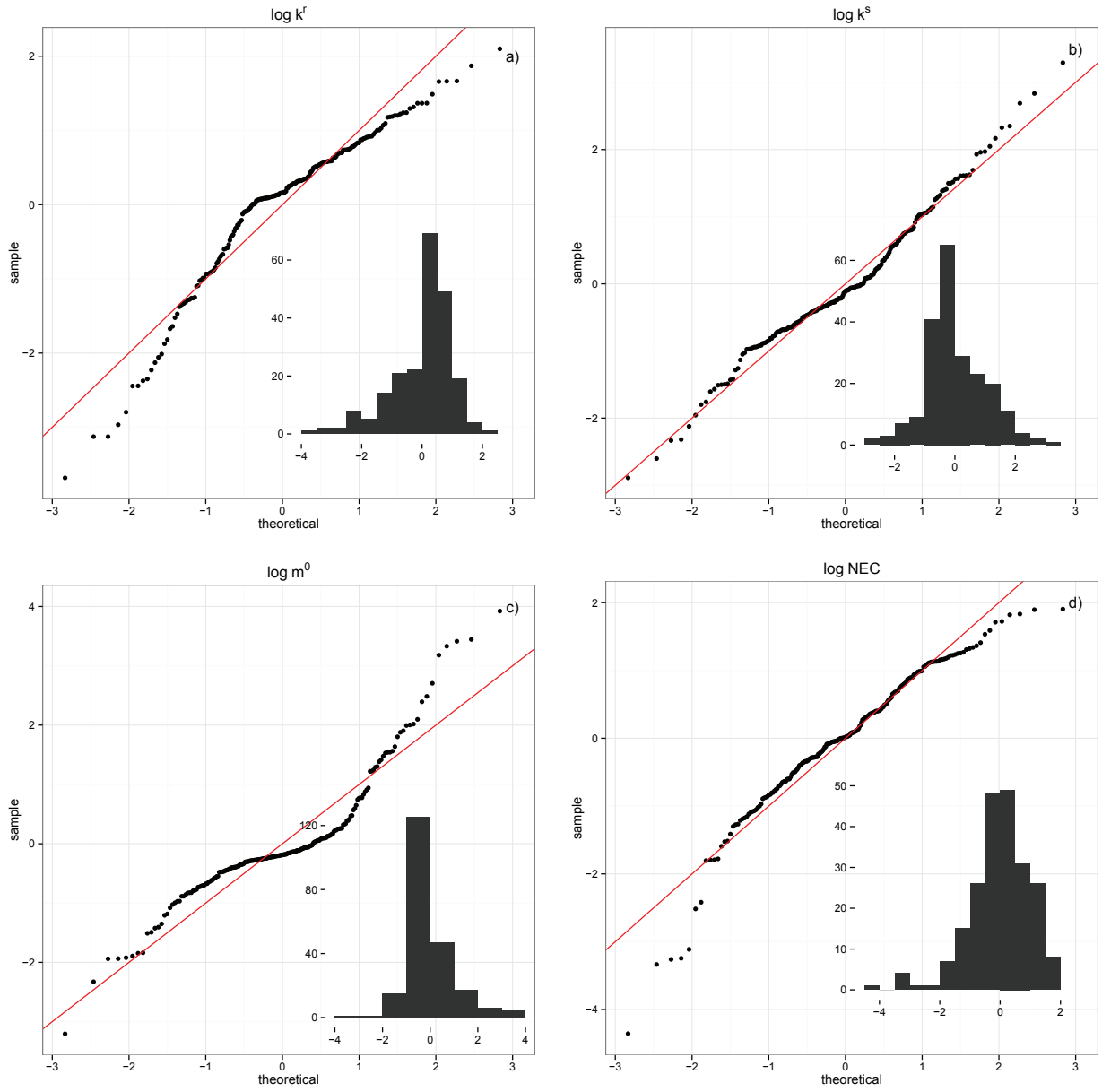



Figure 4.10: Quantile-Quantile plot and histogram of the distribution of each parameter.

### 4.5.4 Practical implications

The time-resolved SSD requires more data than what is usually available in databases such as ECOTOX[eco, ] from US EPA, or etoxBase[RIVM, 2005] from RIVM. Therefore, raw data from the experiments needs to be archived. This is anyway desirable, as current databases store summaries of data of inhomogeneous quality which are treated equivalently in an SSD. With the full raw data it would be possible to account for the precise amount of information collected in the experiments on the sensitivity of the species. As such, our method is also ideally suited to make use of data collected through rapid toxicity testing[Kefford et al., 2005a]. However it is not limited to huge datasets

with hundreds of species as presented here. The model was successfully fitted to a dataset containing only 10 species exposed to triclosan, a organic contaminant used as antibacterial and antifungal. The results were not included in the thesis, but using weakly informative priors[Gelman, 2006] on the scale parameters of the hierarchical model and giving up on modelling correlation, it was possible to use our time-resolved SSD on a dataset of a more common size (the number of species recommended by ECHA is 10, preferably 15[Aldenberg and Rorije, 2013]).

### 4.5.5 Implications for salinity

The hierarchical SSD reduces the discrepancy between the estimated $LC_{50}$ and what is observed in the field, but there remains a difference, hinting at other influences in the field. For example there is approximately a 50% decline in Ephemeroptera, Trichoptera and Plecoptera (EPT) species richness at a salinity of 1000 $\mu S.cm^{-1}$ in south-east Australia[Kefford et al., 2011] while in Central Appalachia, USA 5% of stream invertebrate genera were lost at 295 $\mu S.cm^{-1}$[Cormier et al., 2011]. In comparison, the $HC_5$ based on estimated here was 4407 (95% CI 3310-5634) $\mu S.cm^{-1}$. Dowse et al.[Dowse et al., 2013] compared 72h $LC_{50}$ to sub-lethal chronic endpoints of stream invertebrates exposed to salinity and found the highest Acute to Chronic Ratio (ACR) of 6.1. Such ACR cannot alone explain the difference between loss of taxa in the field and the NEC-based $HC_5$ reported here. Other possible explanations include that only limited sub-lethal salinity sensitivity data are available for salt sensitive stream invertebrates, such as EPT, and these groups may have a higher ACR, that indirect effects of salinity might be propagated via ecological interactions, that there are additional stressors in the field or that salinity just lowers the fitness and affects other life history traits (reproduction, growth, ...), which means that mortality data is not a sufficient endpoint to explain the field observations.

### 4.5.6 Wider application in risk assessment

Our hierarchical TKTD framework could be used for much more than computing an $HC_5$ at each time. Previous work[Ashauer and Wittmer, 2011, Smit et al., 2008] showed that fitted TKTD and Dynamic Energy Budget (DEB) models can be used to predict the response of a community to a realistic exposure scenario. Smit et al.[Smit et al., 2008] estimated five CECs at several dates from a DEBtox model fitted on five species, then modelled the variability of these CEC in a typical two-stages process, ignoring the uncertainty on the estimated CEC. Ashauer et al.[Ashauer and Wittmer, 2011] only modelled the variability on the CEC, essentially fixing the other parameters of the TKTD model at values found for *Gammarus pulex*. The variability on the CEC was then estimated from an ecotoxicological database. The hierarchical model presented here offers the possibility to model the variability on all the parameters of the TKTD, with the interesting addition that the Bayesian framework we used permits a rigorous propagation of the uncertainty on the parameter estimates. Starting from the posterior distribution of the parameters or summarizing it adequately, it would be possible to predict the distribution survival probabilities in a community after an exposition to fluctuating salinity levels.

The distribution of survival probabilities in the community could then be used as a tool for risk management to compute the risk resulting from arbitrary scenarios of exposure, in the same spirit as the forward classical SSD approach[van Straalen, 2010].

## 4.6 Prediction

We give here an example of application for computing the risk due to two different contamination scenarios. We lacked the time to develop a completely realistic example, so this is intended as a schematic proof of concept only.
The idea behind this approach is a variation on the global response defined in subsubsection 3.4.4.1 for diatom communities: since the hierarchical model captures all the information in the raw survival data about the distribution of the TKTD model parameter, it is possible to simulate communities and study their global response. The interesting addition to the global response developed for diatoms is that although the parameters of the TKTD model were estimated with constant concentrations, it is possible to predict a response for scenarios with variable concentrations. The relation between exposure and survival is not trivial. For instance, standard methods use time-weigthed average concentrations[Ashauer and Wittmer, 2011] which assumes that toxicity is determined by the product of exposure time and concentration. Ashauer et al. have compared the effect of multiple concentration peaks during a short time against the effect of peaks more distant in time with similar time-weighted average concentration and found that the simple time-weighted average models failed to predict appropriate mortality in the presence of several concentration peaks [Ashauer et al., 2013]. This was attributed to the absence of a damage recovery process (Equation 4.7) in time-weighted average models which would have allowed the organisms to recover between peaks. Therefore, since observation of the exposure concentration pattern is not enough to predict the effect of a contaminant to a community, simulating a global response from the model is interesting.
As a large proportion of the species in the dataset came from Australia, we set the proof of concept in the context of rising salinity levels in Australia streams. Stream salinity has increased due do land clearing[Nielsen et al., 2003] for instance in River Murray, Victoria, and it reached levels that affect stream communities[Halse et al., 2003]. Management strategies have been implemented[Clarke et al., 2002], consisting in replacing the lost vegetation, establishing watering plans, setting up salt interception schemes, etc. These salt management strategies seem to have been relatively successful in reducing salinity levels at some locations, for instance at Morgan, South Australia, on River Murray (Figure 4.11).
We took monitored salinity data from the Murray Darling Basin Authority (`https://www.waterconnect.sa.gov.au/Systems/SiteInfo/SitePages/Home.aspx?site=A4261110&period=DAILY#SiteSummary`) for the same location: Morgan, South Australia on River Murray. The data was collected daily between 2010 and 2012. Figure 4.11 clearly shows that salinity is now contained at reasonably low levels, and we expect freshwater species to be tolerant of these levels. In order to observe an effect of fluctuating salinity levels on survival, we arbitrarily increased the salinity by multiplying it by 10. This brought the maximum salinity in the dataset to 4800 $\mu S/cm$,

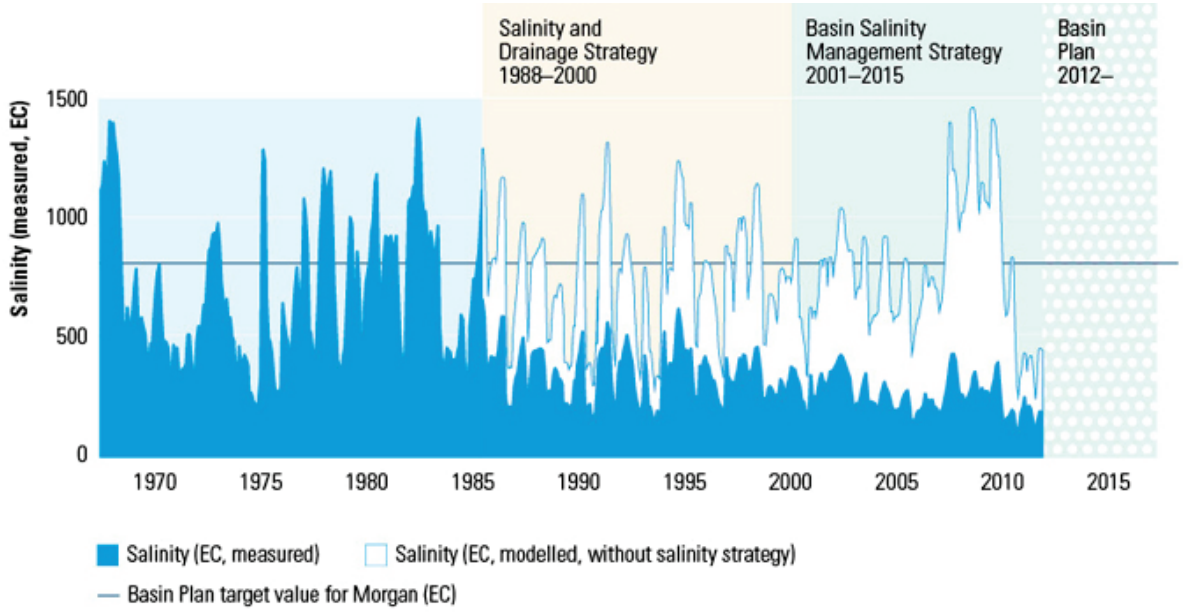



Figure 4.11: River Murray salinity at Morgan over time, and impact of management strategies. Figure borrowed from `http://www.mdba.gov.au/about-mdba/corporate-documents/annual-reports/ar-2012-13/chapter-02`

a little above the $HC_5$ based on NEC (4405 $\mu S/cm[3310 - 5634]$.
We chose to discount the effect of the control mortality parameter because it reflects natural mortality, or mortality resulting from stress unrelated to salinity, and this is not relevant for comparing between exposure scenarios. We computed the survival probability for communities of species by generating communities from the median marginal estimates of the hierarchical model[4]. In the same spirit as the computation of the global response in the case of the diatom hierarchical model, we sampled 100 communities with 500 species each. For each community we computed a global response, and we obtained the uncertainty on that global response by considering the 2.5 and 97.5 percentiles over the communities.
It would be possible to define a global response using the same mathematical formula as for the diatom communities (Equation 3.8):

$$\mathrm{r}_{\mathrm{tot}}(t) = \frac{\sum_{j \in species} S_j(t)}{\mathrm{N}_{\mathrm{species}}} \tag{4.27}$$

where $S_j(t)$ is the response of species $j$ at a time $t$. This $\mathrm{r}_{\mathrm{tot}}(t)$ would be the mean survival probability in the community. However, the mean survival probability of the community does not seem an adequate target of protection, one would want to protect the majority of species in the community. It seemed more interesting to define the global response of a community by the survival probability of the most sensitive species in that

[4]A next step in developing this prediction tool would be to include the full uncertainty from the raw data by using not the median of the marginal posterior distributions but the full joint posterior distribution

community. More precisely, at each time we ranked the species of the community by survival probability and computed the $5^{\text{th}}$ percentile. This percentile is considered as the global response of the community and it can be interpreted as follows: if after $t$ hours the global response is $x\%$, then 95% of the species have a survival rate of more than $x\%$. This global response is very similar in spirit to the $HC_5$: the aim is to protect 95% of the species in the community, and this global response describes the worst effect suffered by the species meant to be protected.

Compared to the global response defined for the diatom communities, this new version has the advantage that it does not require summing the response of each species and assuming that each species contribute equally to the global response. Its focus on the $5^{\text{th}}$ percentile is simply a translation of the 5% threshold used in classical SSD. However, as we noted when we defined the global response for diatom communities, the SSD assumption that all species are considered equally important remains, and no provisions are made for the case where some of the 5% left at risk might play a key role in the community.

We compared the global response thus defined for two scenarios: the salinity level observed at Morgan, South Australia and a scenario corresponding to a schematic remediation strategy succeeding in reducing salinity by 30%, a reduction factor chosen for illustrative purposes. This reduction factor impacts peaks much more than low salinity values, which is consistent with the pattern observed on Figure 4.11. This is a moderate reduction however, as reductions up to 70% can be seen on Figure 4.11.

The time evolution of the global reponse (Figure 4.12) shows a clear difference between the two concentration scenarios. While survival decrease sharply when no remediation strategy is implemented, a 30% reduction of salinity preserves the global response for much longer. Other remediation strategies which would have stronger or more targeted impact could be compared using the same framework.

The time duration of the scenarios is a little disproportionate compared to the time-scale of the experiments (two years versus 96H). Although in principle, the TD parameters are time-invariant and could be used for arbitrary duration, it is likely that such a simple model does not describe accurately survival of species over such a long period. Salinity data in Australia is recorder automatically every 15 minutes in some areas, and it would be possible to use this data for exposure scenarios which would be on the scale of days or weeks. Or maybe salinity does not evolve much on short time scales and the framework developed here would be more applicable for other contaminants. However, we showed that using the posterior distributions of the hierarchical model, is it possible to predict the responses in a community to arbitrary contamination scenarios and we proposed a method to build a global response for the community using the same ideas as the for the $HC_5$ which can be used as an indicator for community protection.

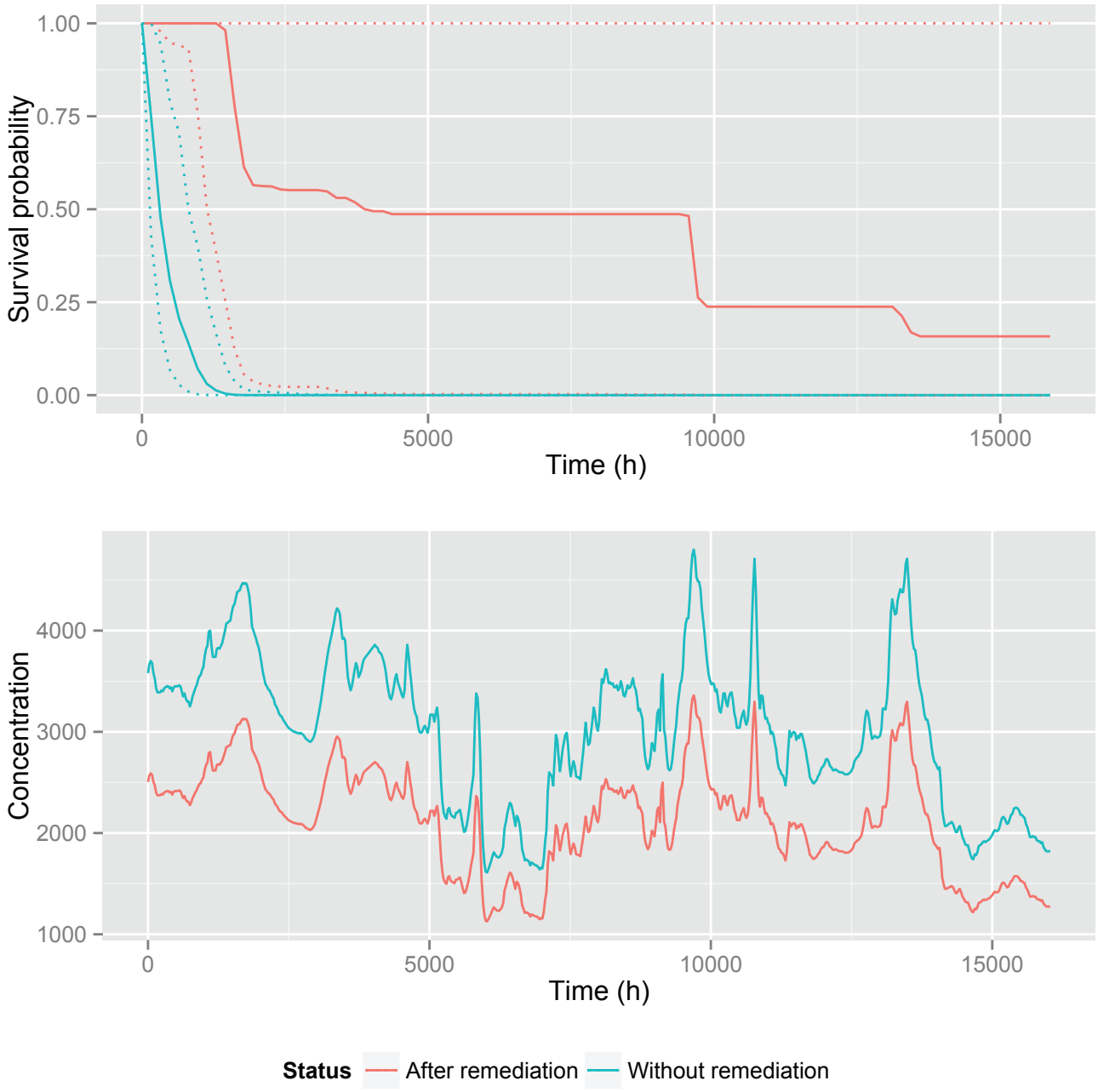



Figure 4.12: Top panel: Comparison of the global response over time for a community of species located at Morgan, South Australia, with or without the implementation of a schematic remediation strategy which induce a 30% reduction of salinity. The dotted lines represent the 95% credible interval. Bottom panel: salinity profile with or without the schematic remediation strategy.

# Conclusion and perspectives

In this thesis, we made several proposals to improve the current SSD method. In the first part of the thesis, we explained how toxicity tests could produce censored toxicity data and we showed a simple method to include it in SSD. We highlighted the adverse effects of the standard practice to discard or transform censored data and provided a web-tool to carry out easily an SSD study on censored data. In the second part of the thesis we noted that in classical SSD, only a summary of the bioassay data is used, discarding valuable information in the process. We showed how to extend the SSD method to include that information using hierarchical modelling of the whole concentration-effect curve. This approach allowed to take uncertainty on the CECs into account in the prediction of the $HC_5$ and to define a $GEC_5$ to characterise the response of the community in terms of the endpoint measured rather than in terms of the impact on biodiversity. The $GEC_5$ could potentially be based on a functional endpoint and make SSD capable of protecting structure as well as function. In the third part of the thesis, we noted that the experiments conducted for an SSD analysis are often followed through time, but that only the result at the end of the experiment is used in classical SSD approaches. This entails interpretability problems for SSD as the predicted $HC_5$ inevitably depends on the duration of the experiment, deteriorating its ecological relevance. Building on the hierarchical model developed for the second part, we proposed to use a dynamical model for survival to describe time-resolved data. This allowed estimating a time-independent NEC for each species which could be used to estimate a limit $HC_5$ valid for any time. As with the hierarchical model of the second part of the thesis, including all the information available in the raw data yielded additional benefits: it allowed including the large uncertainty arising from RTT data, it revealed a considerable variability on all the parameters of the concentration-response model and made it possible to compute a global response for the community. The added value of the mechanistic modelling of the species' response was that this global response could be computed for arbitrary exposure scenarios and used as a tool for comparing salinity remediation strategies.

We took special care not to suggest improvements which would require collecting more data than what is routinely collected in a standard bioassay. Admittedly, there is a need to archive the raw data from the experiments and this type of data is not widely available for the moment. Since this data is already recorded and stored during the experiment, the additional effort required to archive not a summary but the whole raw data would not be so great. There are already several available options for on-line archiving of

scientific datasets such as the Dryad Digital Repository (`http://www.datadryad.org`) which could be used for that purpose.
Including mechanistic effects in SSD is an interesting perspective as it opens the door to developing prediction tools, but models for sublethal effects can be more complex [Kooijman, 2010] than models for survival and hierarchical modelling might prove more difficult. One asset of the TD model we used in this thesis is that it leads to an analytical expression for the likelihood function with constant exposure concentration which makes fitting the hierarchical model almost straightforward. The need to integrate numerically a differential equation to compute the likelihood would make things more complex and require either discretising the equations, using an Ordinary Differential Equation solver (such a functionality is available in Stan) or using methods such as Approximate Bayesian Computation (ABC).
The forecasting tool for concentration scenarios analysis presented in the third part of the thesis must be further developed and validated. The conditions for its applicability to real-world scenarios must be defined properly, notably concerning the time-scale for which it can be relevant. Although the parameters estimated in the TKTD model are time-independent, extrapolation to long time scales must be validated because the TKTD model will not remain valid for species that evolve from larval stage to imago such as mayflies.
The SSD approach has many fundamental weaknesses and they remain in spite of the improvements presented in this thesis. While we focused on improving some of the statistical methodology of SSD, we remained in a fully parametric setting, which required making arbitrary distributional choices. On one hand, this allows the method to remain applicable to smaller datasets and the hierarchical TD model was successfully fitted on a more standard dataset than the salinity dataset, which contained 10 species exposed to triclosan. On the other hand, researchers developing non-parametric SSDs argued with reason about the lack of any mechanistic explanation to underlie these distributional choices. Yet, the parametric setting remains a pragmatic choice when the number of species is too low to allow for other choices.
We did not implement recent developments to account for species non-exchangeability [Craig et al., 2012] and selection bias [Fox, 2015b], and relied on some of the crude assumptions underlying classical SSD. It would be possible in principle to extend the model to account for species non-exchangeability but this relies on borrowing information from other contaminants. A large database of information about other contaminants to tackle this extension is not yet available, but as it might happen it could be possible to implement those in the future, provided the full model does not prove computationally intractable. Another possible development to our approach could be to use threshold distributions as suggested by Van Straalen [van Straalen, 2002]. These distribution forms could be used in the hierarchical models to avoid resorting to the concept of $HC_5$, preferring an $HC_0$, which fits better in the framework of protecting intrinsic value. It is not computationally easy to deal with correlation for every type of distribution though. The common method is to use a gaussian copula, i.e. to generate multivariate normal correlated random variables with a given correlation and to transform them to the target distributions. There is a subtlety in that the correlation of the normal distributed variables is not the same as the target correlation of the variables, because the transfor-

mation does not preserve the Pearson correlation, so the correlation of the multivariate normal variables should be chosen adequately.
Although we also tried to improve on the ecological relevance of SSD, we left some important topics untouched such as the need to account for species interactions, or to deal with mixture of contaminants. We also relied on the assumption that the selected endpoint is ecotoxicologically relevant [Forbes and Calow, 2002], that the assumption that a PNEC based on this endpoint is protective of all the properties of the community. When introducing the $GEC_5$ we tried to move from only protecting a portion of the species to protecting other community features, but we were still dependent on the endpoint tested in the bioassay which was selected not solely for its ecological relevance but also taking into account strong experimental constraints about measurability and repeatability.
The $GEC_5$ for the diatom community was not trivial to define and rested on the assumption that all diatom species contributed equally to the global response. Combining species responses into a global response for the community must be envisioned on a case by case basis regarding the endpoint considered. In the third part of the thesis, we defined a global response in the case of survival data as the a quantile on the response of the most sensitive species in the community. This approach was developed in the same spirit as the $HC_5$ concept: it defined an indicator delimiting between the proportion of species left at risk in SSD and the rest of the community. This approach based on quantiles is more general than the global response and it could also have been applied in the case of the diatom community. However, it is restricted to protecting the structure of the community and cannot be extended to function. The difference between the global response defined in the second and that defined in the third part of the thesis is akin to the difference between mean and median. The mean can be a meaningful combination of individual properties, when those properties are additive[5], but the median (or another quantile) is more widely applicable as it is based only on the ranks.
From a general perspective, we strived to include as much information as possible that was available in the raw data, while many recent developments of SSD focus on including exterior information such as taxonomic information, information from other contaminants, ecological information about the keystone species via weighting, expert judgement, etc. It is certainly one of the most interesting direction for developing the research presented in this thesis. If it can be decided that species are exchangeable within a taxonomic group, for instance, this would justify adding intermediary levels to the hierarchical models based on taxonomic information. Another development could be to base the groups on mode of action of the contaminants, relying on knowledge of the physiology of the tested species. Mechanistic modelling can also be seen as adding external information, relative to the toxicity processes affecting the species, into SSD. Developing the mechanistic models to increase the ecological relevance of SSD and ecological risk assessment in general is the most desirable target for future progress. As environmental protection is much more complex than simply preserving ecosystems in their present states, the only way forward is to move from the static paradigm of SSD to an integrative approach accounting for interactions, evolution, influence of abiotic

[5]biomass of species is a good example of additivity

parameters and ultimately human well-being.

# Bibliography

[eco, ] ECOTOX Database. Technical report, US Agency Environmental Protection, Duluth, Minnesota.

[USE, 1998] (1998). Guidelines for ecological risk assessment. Technical report, US Environmental Protection Agency, Washington, DC.

[min, 2000] (2000). MINITAB statistical software.

[Aagaard et al., 2013] Aagaard, A., Brock, T. C. M., Capri, E., Duquesne, S., Filipic, M., Adriaanse, P. I., and Boesten, J. (2013). Guidance on tiered risk assessment for plant protectionns products for aquatic organisms in edge-of-field surface waters. *EFSA Journal*, 11(7):268.

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.

[Aldenberg and Jaworska, 2000] Aldenberg, T. and Jaworska, J. S. (2000). Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46(1):1–18.

[Aldenberg et al., 2002] Aldenberg, T., Jaworska, J. S., Traas, T. P., and Posthuma, L. (2002). Normal species sensitivity distributions and probabilistic ecological risk assessment. In Posthuma, L., Suter, G. I., and Traas, T. P., editors, *Species sensitivity distributions in ecotoxicology*. Lewis Publishers, Boca Raton, FL.

[Aldenberg and Rorije, 2013] Aldenberg, T. and Rorije, E. (2013). Species Sensitivity Distribution estimation from uncertain (QSAR-based) effects data. *Alternatives to Laboratory Animals*, 41(1):19–31.

[Aldenberg and Slob, 1993] Aldenberg, T. and Slob, W. (1993). Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, 25(1):48–63.

[A.M Verdonck et al., 2001] A.M Verdonck, F., Jaworska, J. S., Thas, O., Vanrolleghem, P. A., and a.M Verdonck, F. (2001). Determining environmental standards using bootstrapping, bayesian and maximum likelihood techniques: a comparative study. *Analytica Chimica Acta*, 446(1-2):427–436.

[ANZECC, 2000] ANZECC (2000). Australian and New Zealand guidelines for fresh and marine water quality. Technical report, Australian and New Zealand Environmental and Conservation Council Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

[Armitage, Peter and Colton, 2005] Armitage, Peter and Colton, T. (2005). Maximum Likelihood. In *Encyclopedia of Biostatistics*, pages 3058–3059. John Wiley & Sons, 2 edition.

[Ashauer et al., 2006] Ashauer, R., Boxall, A., and Brown, C. (2006). Predicting effects on aquatic organisms from fluctuating or pulsed exposure to pesticides. *Environmental toxicology and chemistry / SETAC*, 25(7):1899–1912.

[Ashauer et al., 2010] Ashauer, R., Hintermeister, A., Caravatti, I., Kretschmann, A., and Escher, B. I. (2010). Toxicokinetic and toxicodynamic modeling explains carry-over toxicity from exposure to diazinon by slow organism recovery. *Environmental Science and Technology*, 44(10):3963–3971.

[Ashauer et al., 2013] Ashauer, R., Thorbek, P., Warinton, J. S., Wheeler, J. R., and Maund, S. (2013). A method to predict and understand fish survival under dynamic chemical stress using standard ecotoxicity data. *Environmental Toxicology and Chemistry*, 32(4):954–965.

[Ashauer and Wittmer, 2011] Ashauer, R. and Wittmer, I. (2011). Environmental risk assessment of fluctuating diazinon concentrations in an urban and agricultural catchment using toxicokinetic–toxicodynamic modeling. *Environmental science & technology*, 45(22):9783–9792.

[Baas et al., 2009] Baas, J., Jager, T., and Kooijman, S. (2009). Estimation of no effect concentrations from exposure experiments when values scatter among individuals. *Ecological Modelling*, 220(3):411–418.

[Barnard et al., 2000] Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1312.

[Bates and Watts, 1988] Bates, D. M. and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. John Wiley & Sons, New York.

[Baty et al., 2015] Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J. P., and Delignette-Muller, M. L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *accepted for publication at the Journal of Statistical Software*.

[Bernardo, 1996] Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, 4:111–122.

[Bernardo and Smith, 2000] Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian theory*. John Wiley & Sons, 2nd edition.

[Brain and Cousens, 1989] Brain, P. and Cousens, R. (1989). An equation to describe dose responses where there is stimulation of growth at low doses. *Weed Research*, 29(2):93–96.

[Brooks and Gelman, 1998] Brooks, S. P. S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455.

[Brown et al., 2009] Brown, K., Tomlinson, J., Duncan, J., Hinchcliffe, A., and Palmquist, K. (2009). Critical comparison of available and potential higher tier testing approaches for the risk assessment of plant protection products, considering at least field and semi-field experimental designs, extrapolation from dose-response relationships, and increas. Technical report, EFSA/PPR/2008/01.

[Burnham et al., 2011] Burnham, K. P., Anderson, D. R., and Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1):23–35.

[Calabrese, 2005] Calabrese, E. J. (2005). Paradigm lost, paradigm found: the re-emergence of hormesis as a fundamental dose response model in the toxicological sciences. *Environmental pollution (Barking, Essex : 1987)*, 138(3):379–411.

[Calabrese, 2009] Calabrese, E. J. (2009). Getting the dose–response wrong: why hormesis became marginalized and the threshold model accepted. *Archives of toxicology*, pages 227–247.

[Calabrese and Baldwin, 2003] Calabrese, E. J. and Baldwin, L. a. (2003). Hormesis: the dose-response revolution. *Annual review of pharmacology and toxicology*, 43:175–97.

[Calabrese and Blain, 2011] Calabrese, E. J. and Blain, R. B. (2011). The hormesis database: the occurrence of hormetic dose responses in the toxicological literature. *Regulatory toxicology and pharmacology : RTP*, 61(1):73–81.

[Campbell et al., 2000] Campbell, E., Palmer, M., Shao, Q., Warne, M. S. J., and Wilson, D. (2000). BurrliOZ: A computer program for calculating toxicant trigger values for the ANZECC and ARMCANZ water quality guidelines.

[Casella and Berger, 2002] Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2 edition.

[CCME, 2007] CCME (2007). A protocol for the derivation of water quality guidelines for the protection of aquatic life. In *Canadian Environmental Quality Guidelines*. Canadian Council of Ministers of the Environment, Winnipeg.

[Cedergreen et al., 2009] Cedergreen, N., Ritz, C., and Streibig, J. C. (2009). Improved empirical models describing hormesis. *Environmental Toxicology and Chemistry Toxicology and Chemistry*, 24(12):3166–3172.

[Chapman et al., 1996] Chapman, P. M., Caldwell, R. S., and Chapman, P. F. (1996). A warning: NOECs are inappropriate for regulatory use. *Environmental Toxicology and Chemistry*, 15(2):77–79.

[Chen, 2004] Chen, L. (2004). A conservative, nonparametric estimator for the 5th percentile of the species sensitivity distributions. *Journal of Statistical Planning and Inference*, 123(2):243–258.

[Ciffroy et al., 2012] Ciffroy, P., Keller, M., and Pasanisi, A. (2012). Estimating hazardous concentrations by an informative bayesian approach. *Environmental toxicology and chemistry / SETAC*, (October).

[Clarke et al., 2002] Clarke, C. J., George, R. J., Bell, R. W., and Hatton, T. J. (2002). Dryland salinity in south-western Australia: Its origins, remedies, and future research directions. *Australian Journal of Soil Research*, 40(1):93–113.

[Cormier et al., 2011] Cormier, S. M., Suter II, G. W., Yuan, L. L., and Zheng, L. (2011). A field-based aquatic life benchmark for conductivity in central Appalachian streams. Technical report, EPA/600/R-10.

[Coutris et al., 2011] Coutris, C., Merlina, G., Silvestre, J., Pinelli, E., and Elger, A. (2011). Can we predict community-wide effects of herbicides from toxicity tests on macrophyte species? *Aquatic Toxicology*, 101(1):49–56.

[Craig, 2013] Craig, P. S. (2013). Exploring novel ways of using species sensitivity distributions to establish PNECs for industrial chemicals: Final report to Project Steering Group. Technical report.

[Craig et al., 2012] Craig, P. S., Hickey, G. L., Luttik, R., and Hart, A. (2012). Species non-exchangeability in probabilistic ecotoxicological risk assessment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(1):243–262.

[Cropper, 1993] Cropper, A. (1993). Convention on Biological Diversity. *Environmental Conservation*, 20(04):364.

[Daniels and Kass, 1999] Daniels, M. J. and Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94(448):1254–1263.

[Davis, 2011] Davis, M. (2011). Don ' t judge species on their origins. *Nature*, 474(9):153–154.

[De Laender et al., 2008] De Laender, F., De Schamphelaere, K. a. C., Vanrolleghem, P. A., and Janssen, C. R. (2008). Is ecosystem structure the target of concern in ecological effect assessments? *Water research*, 42(10-11):2395–402.

[Delignette-Muller and Dutang, 2015] Delignette-Muller, M. L. and Dutang, C. (2015). {fitdistrplus}: An {R} Package for Fitting Distributions. *Journal of Statistical Software*, 64(4):1–34.

[Delignette-Muller et al., 2011] Delignette-Muller, M. L., Forfait, C., Billoir, E., and Charles, S. (2011). A new perspective on the Dunnett procedure: Filling the gap between NOEC/LOEC and ECx concepts. *Environmental Toxicology and Chemistry*, 30(12):2888–2891.

[Delignette-Muller et al., 2013] Delignette-Muller, M. L., Pouillot, R., Denis, J.-B., and Dutang, C. (2013). fitdistrplus: help to fit of a parametric distribution to non-censored or censored data.

[Dixon, 2007] Dixon, W. J. (2007). The use of Probability Bounds Analysis for Characterising and Propagating Uncertainty in Species Sensitivity Distributions. Technical Report 163.

[Dowse et al., 2013] Dowse, R., Tang, D., Palmer, C. G., and Kefford, B. J. (2013). Risk assessment using the species sensitivity distribution method: data quality versus data quantity. *Environmental Toxicology and Chemistry*, 32(6):1360–1369.

[Duboudin et al., 2004a] Duboudin, C., Ciffroy, P., and Magaud, H. (2004a). Acute-to-chronic species sensitivity distribution extrapolation. *Environmental toxicology and chemistry / SETAC*, 23(7):1774–1785.

[Duboudin et al., 2004b] Duboudin, C., Ciffroy, P., and Magaud, H. (2004b). Effects of data manipulation and statistical methods on species sensitivity distributions. *Environmental Toxicology and Chemistry*, 23(2):489–499.

[ECHA, 2008] ECHA (2008). Guidance on information requirements and chemical safety assessment. Technical Report May, Helsinki.

[ECHA, 2012] ECHA (2012). Uncertainty analysis. Technical report.

[Efron and Tibshirani, 1993] Efron, B. and Tibshirani, R. (1993). An introduction to the bootstrap. In *Monographs on Statistics and Applied Probability*, volume 57. CRC press.

[Forbes and Calow, 2002] Forbes, V. E. and Calow, P. (2002). Species sensitivity distributions revisited: a critical appraisal. *Human and Ecological Risk Assessment*, 8(3):473–492.

[Forfait-Dubuc et al., 2012] Forfait-Dubuc, C., Charles, S., Billoir, E., and Delignette-Muller, M. L. (2012). Survival data analyses in ecotoxicology: critical effect concentrations, methods and models. What should we use? *Ecotoxicology*, 21(4):1–12.

[Fox, 2008] Fox, D. R. (2008). NECS, NOECS and the EC. *Australasian Journal of Ecotoxicology*, 14:7–9.

[Fox, 2010] Fox, D. R. (2010). A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicology and environmental safety*, 73(2):123–31.

[Fox, 2015a] Fox, D. R. (2015a). More noise does not mean more precision: A review of Aldenberg and Rorije (2013). *Alternatives to laboratory animals: ATLA*, 43(4):241–249.

[Fox, 2015b] Fox, D. R. (2015b). Selection bias correction for Species Sensitivity Distribution modelling and hazardous concentration estimation. *Environmental Toxicology and Chemistry*, 3098.

[Fox and Billoir, 2013] Fox, D. R. and Billoir, E. (2013). Time-dependent species sensitivity distributions. *Environmental toxicology and chemistry / SETAC*, 32(2):378–83.

[Fox et al., 2012] Fox, D. R., Billoir, E., Charles, S., Delignette-Muller, M. L., and Lopes, C. (2012). What to do with NOECS/NOELS prohibition or innovation? *Integrated environmental assessment and management*, 8(4):764–766.

[Gelman, 2006] Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533.

[Gelman, 2014] Gelman, A. (2014). RStan : the R interface to Stan.

[Gelman et al., 2013] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

[Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Taylor & Francis, third edit edition.

[Gelman and Hill, 2007] Gelman, A. and Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Policy Analysis*, pages 1–651.

[Gelman and Rubin, 1992] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple simulations. *Statistical Science*, 7(4):457–511.

[Grist and Leung, 2002] Grist, E. P. M. and Leung, K. M. Y. (2002). Better bootstrap estimation of hazardous concentration thresholds for aquatic assemblages. *Environmental Toxicology and Chemistry*, 21(7):1515–1524.

[Grist et al., 2006] Grist, E. P. M., O'Hagan, A., Crane, M., Sorokin, N., Sims, I., and Whitehouse, P. (2006). Bayesian and time-independent species sensitivity distributions for risk assessment of chemicals. *Environmental science & technology*, 40(1):395–401.

[Halse et al., 2003] Halse, S. a., Ruprecht, J. K., and Pinder, a. M. (2003). Salinisation and prospects for biodiversity in rivers and wetlands of south-west Western Australia. *Australian Journal of Botany*, 51(6):673–688.

[Hartigan and Hartigan, 1985] Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, pages 70–84.

[Hayashi and Kashiwagi, 2010] Hayashi, T. I. and Kashiwagi, N. (2010). A Bayesian Method for Deriving Species-Sensitivity Distributions: Selecting the Best-Fit Tolerance Distributions of Taxonomic Groups. *Human and Ecological Risk Assessment*, 16(2):251–263.

[He et al., 2014] He, W., Qin, N., Kong, X., Liu, W., Wu, W., He, Q., Yang, C., Jiang, Y., Wang, Q., Yang, B., and Xu, F. (2014). Ecological risk assessment and priority setting for typical toxic pollutants in the water from Beijing-Tianjin-Bohai area using Bayesian matbugs calculator (BMC). *Ecological Indicators*, 45:209–218.

[Helsel, 2005] Helsel, D. R. (2005). *Nondetects and data analysis. Statistics for censored environmental data.* Wiley-Interscience.

[Helsel, 2006] Helsel, D. R. (2006). Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere*, 65(11):2434–9.

[Hickey, 2010] Hickey, G. L. (2010). Ecotoxicological risk assessment: developments in PNEC estimation. (February).

[Hickey and Craig, 2012] Hickey, G. L. and Craig, P. S. (2012). Competing statistical methods for the fitting of normal species sensitivity distributions: recommendations for practitioners. *Risk Analysis*, 32(7):1232–43.

[Hickey et al., 2009] Hickey, G. L., Craig, P. S., and Hart, A. (2009). On the application of loss functions in determining assessment factors for ecological risk. *Ecotoxicology and Environmental Safety*, 72(2):293–300.

[Hickey et al., 2012] Hickey, G. L., Craig, P. S., Luttik, R., and de Zwart, D. (2012). On the quantification of intertest variability in ecotoxicity data with application to species sensitivity distributions. *Environmental toxicology and chemistry / SETAC*, 31(8):1903–10.

[Hickey and Kefford, 2008] Hickey, G. L. and Kefford, B. J. (2008). Making species salinity sensitivity distributions reflective of naturally occurring communities: using rapid testing and Bayesian statistics. *Environmental Toxicology and Chemistry*, 27(11):2403–2411.

[Hoffman and Gelman, 2011] Hoffman, M. D. and Gelman, A. (2011). No Title.

[Hooper et al., 2005] Hooper, D., Chapin III, F., and Ewel, J. (2005). Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, 75(1):3–35.

[Jager et al., 2011] Jager, T., Albert, C., Preuss, T. G., and Ashauer, R. (2011). General unified threshold model of survival - a toxicokinetic-toxicodynamic framework for ecotoxicology. *Environmental science & technology*, 45(7):2529–40.

[Jagoe and Newman, 1997] Jagoe, R. H. and Newman, M. C. (1997). Bootstrap estimation of community NOEC values. *Ecotoxicology*, 6(5):293–306.

[Jaynes, 2003] Jaynes, E. T. (2003). Preface. In *Probability theory: the logic of science*, pages xxv–xxvi. Cambridge university press.

[Kaplan and Meier, 1958] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

[Kefford et al., 2012a] Kefford, B. J., Hickey, G. L., Gasith, A., Ben-David, E., Dunlop, J. E., Palmer, C. G., Allan, K., Choy, S. C., and Piscart, C. (2012a). Global scale variation in the salinity sensitivity of riverine macroinvertebrates: eastern Australia, France, Israel and South Africa. *PloS one*, 7(5):e35224.

[Kefford et al., 2011] Kefford, B. J., Marchant, R., Schäfer, R. B., Metzeling, L., Dunlop, J. E., Choy, S. C., and Goonan, P. (2011). The definition of species richness used by species sensitivity distributions approximates observed effects of salinity on stream macroinvertebrates. *Environmental pollution (Barking, Essex : 1987)*, 159(1):302–10.

[Kefford et al., 2006] Kefford, B. J., Nugegoda, D., Metzeling, L., and Fields, E. J. (2006). Validating species sensitivity distributions using salinity tolerance of riverine macroinvertebrates in the southern Murray-Darling Basin (Victoria, Australia). *Canadian Journal of Fisheries and Aquatic Sciences*, 1877:1865–1877.

[Kefford et al., 2005a] Kefford, B. J., Palmer, C. G., Jooste, S., Warne, M. S. J., and Nugegoda, D. (2005a). What is meant by '95% of species'? An argument for the inclusion of rapid tolerance testing.

[Kefford et al., 2005b] Kefford, B. J., Palmer, C. G., and Nugegoda, D. (2005b). Relative salinity tolerance of freshwater macroinvertebrates from the south-east Eastern Cape, South Africa compared with the Barwon Catchment, Victoria, Australia. *Marine and Freshwater Research*, 56:163–171.

[Kefford et al., 2003] Kefford, B. J., Papas, P. J., and Nugegoda, D. (2003). Relative salinity tolerance of macroinvertebrates from the Barwon River, Victoria, Australia. *Marine and Freshwater Research*, 54(6):755–765.

[Kefford et al., 2012b] Kefford, B. J., Schäfer, R. B., and Metzeling, L. (2012b). Risk assessment of salinity and turbidity in Victoria (Australia) to stream insects' community structure does not always protect functional traits. *The Science of the total environment*, 415:61–8.

[Kefford et al., 2008] Kefford, B. J., Zalizniak, L., Warne, M. S. J., and Nugegoda, D. (2008). Is the integration of hormesis and essentiality into ecotoxicology now opening Pandora's Box? *Environmental pollution (Barking, Essex : 1987)*, 151(3):516–23.

[Klein and Goel, 2013] Klein, J. P. and Goel, P. (2013). *Survival analysis: state of the art*, volume 211. Springer Science & Business Media.

[Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

[Kon Kam King et al., 2015] Kon Kam King, G., Larras, F., Charles, S., and Delignette-muller, M. L. (2015). Hierarchical modelling of species sensitivity distribution: development and application to the case of diatoms exposed to several herbicides. *Ecotoxicology and Environmental Safety*, 114:212–221.

[Kon Kam King et al., 2014] Kon Kam King, G., Veber, P., Charles, S., and Delignette-Muller, M. L. (2014). MOSAIC_SSD: a new web tool for species sensitivity distribution to include censored data by maximum likelihood. *Environmental toxicology and chemistry / SETAC*, 33(9):2133–9.

[Kooijman, 1987] Kooijman, S. A. L. M. (1987). A safety factor for LC 50 values allowing for differences in sensitivity among species. *Water Research*, 21(3):269–276.

[Kooijman, 2010] Kooijman, S. A. L. M. (2010). *Dynamic energy budget theory for metabolic organisation*. Cambridge university press.

[Kooijman and Bedaux, 1996] Kooijman, S. A. L. M. and Bedaux, J. J. M. (1996). *The analysis of aquatic toxicity data*. VU University press Amsterdam, Amsterdam.

[Kooijman et al., 1996] Kooijman, S. A. L. M., Bedaux, J. J. M., and Slob, W. (1996). No-effect concentration as a basis for ecological risk assessment. *Risk analysis : an official publication of the Society for Risk Analysis*, 16(4):445–7.

[Koyama, 1996] Koyama, J. (1996). Vertebral deformity susceptibilities of marine fishes exposed to herbicide. *Bulletin of Environmental Contamination and Toxicology*, 56(4):655–62.

[Kroll and Stedinger, 1996] Kroll, C. and Stedinger, J. (1996). Estimation of moments and quantiles using censored data. *Water Resources Research*, 32(4):1005–1012.

[Larras et al., 2012] Larras, F., Bouchez, A., Rimet, F., and Montuelle, B. (2012). Using Bioassays and Species Sensitivity Distributions to Assess Herbicide Toxicity towards Benthic Diatoms. *PloS one*, 7(8):e44458.

[Lee, 2013] Lee, L. (2013). NADA: Nondetects And Data Analysis for environmental data.

[Lewandowski et al., 2009] Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

[Liu et al., 2014] Liu, Y., Wu, F., Mu, Y., Feng, C., Fang, Y., Chen, L., and Giesy, J. P. (2014). Setting Water Quality Criteria in China: Approaches for Developing Species Sensitivity Distributions for Metals and Metalloids. In *Reviews of Environmental Contamination and Toxicology volume*, pages 35–57. Springer.

[Lunn et al., 2000] Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.

[Mayer et al., 2002] Mayer, F. L., Ellersieck, M. R., Krause, G. F., Sun, K., Lee, G., and Buckler, D. R. (2002). *Time-concentration effect models in predicting chronic toxicity from acute toxicity data.* Boca Raton, FL.

[McGill et al., 1978] McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

[Moore et al., 2010] Moore, D., Warren-Hicks, W., Qian, S., Fairbrother, A., Aldenberg, T., Barry, T., Luttik, R., and Ratte, H. (2010). Uncertainty analysis using classical and bayesian hierarchical models. In Warren-Hicks, W. and Hart, A., editors, *Application of Uncertainty Analysis to Ecological Risk of Pesticides*, pages 134–141. CRC press, Pensacola (Florida), US.

[Newman et al., 2000] Newman, M. C., Ownby, D. R., Mézin, L. C. A., Powell, D. C., Christensen, T. R. L., Lerberg, S. B., and Anderson, B.-A. (2000). Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. *Environmental Toxicology and Chemistry*, 19(2):508–515.

[Nielsen et al., 2003] Nielsen, D. L., Brock, M. a., Rees, G. N., and Baldwin, D. S. (2003). Effects of increasing salinity on freshwater ecosystems in Australia. *Australian Journal of Botany*, 51(6):655–665.

[Nugegoda and Kibria, 2013] Nugegoda, D. and Kibria, G. (2013). Water Quality Guidelines for the Protection of Aquatic Ecosystems. In *Encyclopedia of Aquatic Ecotoxicology*, pages 1177–1196. Springer.

[O'Hagan et al., 2005] O'Hagan, A., Crane, M., Grist, E. P. M., and Whitehouse, P. (2005). Estimating species sensitivity distributions with the aid of expert judgements. *Unpublished manuscript*, pages 1–12.

[Pesce et al., 2009] Pesce, S., Batisson, I., Bardot, C., Fajon, C., Portelli, C., Montuelle, B., and Bohatier, J. (2009). Response of spring and summer riverine microbial communities following glyphosate exposure. *Ecotoxicology and environmental safety*, 72(7):1905–1912.

[Pires et al., 2002] Pires, A. M., Branco, J. A., Picado, A., and Mendonça, E. (2002). Models for the estimation of a 'no effect concentration'. *Environmetrics*, 13(1):15–27.

[Plummer, 2003] Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pages 20–22.

[Plummer, 2013] Plummer, M. (2013). JAGS Version 3.4.0 user manual.

[Posthuma et al., 2010] Posthuma, L., Suter II, G. W., and Traas, T. P. (2010). *Species sensitivity distributions in ecotoxicology.* CRC press.

[Power and McCarty, 1997] Power, M. and McCarty, L. S. (1997). Fallacies in Ecological Risk Assessment Practices. *Environmental Science and Technology*, 31(8):370A–375A.

[R Core Team, 2013] R Core Team (2013). R: A Language and Environment for Statistical Computing.

[Ritz, 2010] Ritz, C. (2010). Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry*, 29(1):220–229.

[Ritz and Streibig, 2005] Ritz, C. and Streibig, J. C. (2005). Bioassay analysis using R. *Journal of Statistical Software*, 12(5):1–22.

[RIVM, 2005] RIVM (2005). The RIVM etoxBase. Database for Ecotoxicological Risk Assessment.

[Robert, 2007] Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

[Rodney et al., 2008] Rodney, S., Moore, D., and Teed, R. (2008). SSD Master Version 2.0. Determination of hazardous concentrations with species sensitivity distributions. Intrinsik Inc.

[Roux et al., 1996] Roux, D. J., Jooste, S., and MacKay, H. M. (1996). Substance-specific water quality criteria for the protection of South African freshwater ecosystems: Methods for derivation and initial results for some inorganic toxic. *South African Journal of Science*, 92:198–206.

[Shao, 2000] Shao, Q. (2000). Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. *Environmetrics*, 11(5):583–595.

[Shapiro and Wilk, 1965] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, pages 591–611.

[Shaw-Allen and Suter II, ] Shaw-Allen, P. and Suter II, G. W. CADDIS SSD Generator.

[Singh et al., 2010] Singh, M., Lodha, P., and Singh, G. P. (2010). Seasonal diatom variations with reference to physico-chemical properties of water of Mansagar lake of Jaipur, Rajasthan. *Research Journal of Agricultural Sciences*, 1(4):451–457.

[Smit and Ebbens, 2008] Smit, M. G. D. and Ebbens, E. (2008). Time and concentration dependency in the potentially affected fraction of species: the case of hydrogen peroxide treatment of ballast water. *Environmental Toxicology and Chemistry*, 27(3):746–753.

[Smit et al., 2008] Smit, M. G. D., Holthaus, K. I. E., Trannum, H. C., Neff, J. M., Kjeilen-Eilertsen, G., Jak, R. G., Singsaas, I., Huijbregts, M. a. J., and Hendriks, J. (2008). Species sensitivity distributions for suspended clays, sediment burial, and grain size change in the marine environment. *Environmental toxicology and chemistry / SETAC*, 27(4):1006–12.

[Spiegelhalter et al., 2002] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

[Srivastava and Vellend, 2005] Srivastava, D. S. and Vellend, M. (2005). BIODIVERSITY-ECOSYSTEM FUNCTION RESEARCH: Is It Relevant to Conservation? *Annual Review of Ecology, Evolution, and Systematics*, 36(1):267–294.

[Stan Development Team, 2014] Stan Development Team (2014). Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0.

[Therneau, 2014] Therneau, T. M. (2014). A Package for Survival Analysis in S.

[Tilman et al., 2001] Tilman, D., Reich, P. B., Knops, J., Wedin, D., Mielke, T., and Lehman, C. (2001). Diversity and productivity in a long-term grassland experiment. *Science (New York, N.Y.)*, 294(5543):843–845.

[Tizzoni et al., 2014] Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Computational Biology*, 10(7).

[Turnbull, 1976] Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology.*, 38(3):290–295.

[van der Hoeven, 1997] van der Hoeven, N. (1997). How to measure no effect. Part III: Statistical aspects of NOEC, ECx and NEC estimates. *Environmetrics*, 8(April 1995):255–261.

[van der Hoeven, 2001] van der Hoeven, N. (2001). Estimating the 5-percentile of the species sensitivity distributions without any assumptions about the distribution. *Ecotoxicology*, 10(1):25–34.

[van der Hoeven, 2004] van der Hoeven, N. (2004). Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheoretica*, 52:201–217.

[van Straalen, 2002] van Straalen, N. M. (2002). Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. *Environmental Toxicology and Pharmacology*, 11(3-4):167–72.

[van Straalen, 2010] van Straalen, N. M. (2010). Theory of Ecological Risk Assessment based on SSDs. In Posthuma, L., Suter II, G. W., and Traas, T. P., editors, *Species Sensitivity Distributions in Ecotoxicology*, chapter Theory of, pages 37–48. CRC press, Boca Raton, FL.

[van Vlaardingen et al., 2004] van Vlaardingen, P. L. A., Traas, T. P., Wintersen, A. M., and Aldenberg, T. (2004). ETX 2.0. pages 1–68.

[Von Mises, 1964] Von Mises, R. (1964). *Mathematical theory of probability and statistics*. New York, academic p edition.

[Wagner and Lokke, 1991] Wagner, C. and Lokke, H. (1991). Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Research*, 25(10):1237–1242.

[Wang et al., 2008] Wang, B., Yu, G., Huang, J., and Hu, H. (2008). Development of species sensitivity distributions and estimation of HC(5) of organochlorine pesticides with five statistical approaches. *Ecotoxicology*, 17(8):716–24.

[Wang et al., 2015] Wang, Y., Wu, F., Giesy, J. P., Feng, C., Liu, Y., Qin, N., and Zhao, Y. (2015). Non-parametric kernel density estimation of species sensitivity distributions in developing water quality criteria of metals. *Environmental Science and Pollution Research*.

[Warne et al., 2013] Warne, M. S. J., Batley, G. E., Braga, O., Chapman, J. C., Fox, D. R., Hickey, G. L., Stauber, J. L., and Van Dam, R. (2013). Revisions to the derivation of the Australian and New Zealand guidelines for toxicants in fresh and marine waters. *Environmental Science and Pollution Research*, pages 1–10.

[Warne and Van Dam, 2008] Warne, M. S. J. and Van Dam, R. (2008). NOEC and LOEC data should no longer be generated or used. *Australasian Journal of Ecotoxicology*, 14:1–5.

[Watson et al., 2003] Watson, R. T., Rosswall, T., Steiner, A., Töpfer, K., Arico, S., and Bridgewater, P. (2003). *Ecosystems and Human Well-being: a Framework for Assessment*. ISLAND PRESS, Washington DC.

[Wheeler et al., 2002] Wheeler, J. R., Grist, E. P. M., Leung, K. M. Y., Morritt, D., and Crane, M. (2002). Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin*, 45(1-12):192–202.

[Xing et al., 2014] Xing, L., Liu, H., Zhang, X., Hecker, M., Giesy, J. P., and Yu, H. (2014). A comparison of statistical methods for deriving freshwater quality criteria for the protection of aquatic organisms. *Environmental Science and Pollution Research*, 21(1):159–167.

[Xu et al., 2015] Xu, F.-L., Li, Y.-L., Wang, Y., He, W., Kong, X.-Z., Qin, N., Liu, W.-X., Wu, W.-J., and Jorgensen, S. E. (2015). Key issues for the development and application of the species sensitivity distribution (SSD) model for ecological risk assessment. *Ecological Indicators*, 54:227–237.

[Zajdlik et al., 2009] Zajdlik, B. A., Dixon, D. G., and Stephenson, G. (2009). Estimating Water Quality Guidelines for Environmental Contaminants Using Multimodal Species Sensitivity Distributions: A Case Study with Atrazine. *Human and Ecological Risk Assessment*, 15(3):554–564.

# Publications

## Publication 1:

Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Computational Biology*, 10(7)

## Publication 2:

Kon Kam King, G., Veber, P., Charles, S., and Delignette-Muller, M. L. (2014). MOSAIC_SSD: a new web tool for species sensitivity distribution to include censored data by maximum likelihood. *Environmental toxicology and chemistry / SETAC*, 33(9):2133–9

## Publication 3:

Kon Kam King, G., Larras, F., Charles, S., and Delignette-muller, M. L. (2015). Hierarchical modelling of species sensitivity distribution: development and application to the case of diatoms exposed to several herbicides. *Ecotoxicology and Environmental Safety*, 114:212–221

## Publication 4:

Kon Kam King, G., Delignette-Muller, M.L., Kefford, B., Piscart, C., Charles, S. (2015). Constructing a time-resolved Species Sensitivity Distribution using a hierarchical Toxico-Kinetic Toxico-Dynamic model. *Environmental Science & Technology* (accepted with major revisions)

Appendix A

# Species Sensitivity Distribution across the world

Here is reproduced the detailed comparison of the water quality guidelines on SSD of several environmental agencies across the world written in [Nugegoda and Kibria, 2013].

**Water Quality Guidelines for the Protection of Aquatic Ecosystems, Table 1** A comparison of world major water quality guidelines for the protection of aquatic life/aquatic ecosystems protection and their derivation methodologies. The information is based on [a]ANZECC and ARMCANZ (2000); [b]CCME (1999); [c]ECB (2003), [d]RIVM (2001), [e]US EPA (1985), [f]US EPA (2003), TenBrook et al. 2009)

| Criteria description | | | | Criteria derivation methodologies | | |
|---|---|---|---|---|---|---|
| Country and year | Guidelines name | Criterion | Criterion description | Assessment factor (AF) method | Species sensitivity distribution (SSD) method | Criteria component |
| Australia and New Zealand, 2000[a] | Australia and New Zealand guidelines for fresh and marine water quality | High reliability trigger values (HRTV) | Derived from > 1 multispecies or > 5 single-species chronic values; exceedance triggers further investigation; not a mandatory standard | – Minimum number of values required: 1<br>– Minimum number of taxa required: 1 | – Burr family/best fit<br>– Minimum number of values required: 5<br>– Minimum number of taxa required: 5<br>– Uncertainty quantified: yes<br>– All data used: yes | Chronic;Magnitude; Bioaccumulation; Bioavailability; Water quality |
| | | Medium reliability trigger values (MRTV) | Derived from > 5 acute data; exceedance triggers further investigation; not a mandatory standard | | | |
| | | Low reliability trigger values (MRTV) | Derived from <5 acute or chronic values; not used as a guideline | | | |
| Canada, 1999[b] | A protocol for the derivation of water quality guidelines for the protection of aquatic life | Guidelines | Single maximum which is not to be exceeded | – Minimum number of values required: 1<br>– Minimum number of taxa required: 1 | – several models tested/best fit<br>– minimum number of values required: not specified as long as best fit and taxa requirements are met<br>– minimum number of taxa required: 7 for freshwater, 6 for marine water | Chronic; Magnitude |

(*continued*)

**Water Quality Guidelines for the Protection of Aquatic Ecosystems, Table 1** (continued)

| Criteria description | | | | Criteria derivation methodologies | | |
|---|---|---|---|---|---|---|
| Country and year | Guidelines name | Criterion | Criterion description | Assessment factor (AF) method | Species sensitivity distribution (SSD) method | Criteria component |
| | | | | | – uncertainty quantified: yes<br>– separate SSDs for short-term exposure and long-term exposure. | |
| European Union, 2003[c] | Technical guidance document on risk assessment, Part II. Environmental risk assessment | Predicted no effect concentration (PNEC) | Used in risk assessment | – Minimum number of values required: 1<br>– Minimum number of taxa required: 1 | – Minimum number of values required: 10<br>– Minimum number of taxa required: 8<br>– Uncertainty quantified: yes<br>– All data used: yes | Chronic;Magnitude; Bioaccumulation; |
| Netherlands, 2001[d] | Guidance document on deriving environmental risk limits in the Netherlands | Negligible concentration (NC) | Used to set environmental quality standards (EQS); EQS may or may not be legally binding | – Minimum number of values required: 1<br>– Minimum number of taxa required: 1 | – Log normal<br>– Minimum number of values required: 4<br>– Minimum number of taxa required: 4<br>– Uncertainty quantified: yes<br>– All data used: yes | Chronic;Magnitude; Bioaccumulation; Threatened and endangered species (TES) |
| | | Maximum permissible concentration (MPC) | Used to set EQS; EQS may or may not be legally binding | | | |
| | | Ecosystem serious risk concentration ($SRC_{ECO}$) | Used to set EQS; EQS may or may not be legally binding | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| USA, 1985[e] | Guidelines for deriving numerical water quality criteria for the protection of aquatic organisms and their uses | CMC: criterion maximum concentration<br>CCC: criterion continuous concentration | Used for setting water quality standards, setting discharge limits, and other regulatory programs; for protection from short-term exposure | – Minimum number of values required: 6-9<br>– Minimum number of taxa required: 5 | – Log triangular<br>– Minimum number of values required: 8 | Acute; Chronic; Magnitude; Duration; Frequency; Bioaccumulation; Mixtures; Bioavailability; Water quality; Threatened and endangered species (TES) |
| USA, 2003[f] | Water quality guidance for the Great Lakes system | Tier I criterion maximum concentration (CMC)<br>Tier I criterion continuous concentration (CCC)<br>Tier II criterion maximum concentration (CMC)<br>Tier II criterion continuous concentration (CCC) | Adopted into water quality standards or used to implement narrative criteria; for protection of short-term exposure<br>Adopted into water quality standards or used to implement narrative criteria; for protection of long-term exposure<br>Used only in implementation of narrative criteria; for protection from short-term exposure<br>Used only in implementation of narrative criteria; for protection from long-term exposure | – Minimum number of values required: 1<br>– Minimum number of taxa required: 1 | – Log triangular<br>– Burr family/ best fit<br>– Minimum number of values required: 8<br>– Minimum number of taxa required: 8 | Acute; Chronic; Magnitude; Duration; Frequency; Bioavailability; Water quality; Threatened and endangered species (TES) |

Appendix B

# The Bayesian paradigm

Bayesian approaches to the SSD have been the occasion of many developments recently. The Bayesian paradigm offers a flexible framework for embedding multiple sources of information in classical SSD. The Bayesian paradigm differs from the frequentist in that it considers the data as fixed and the parameters as distributed, whereas the frequentist paradigm considers that the parameters are fixed and that the data is one realisation of an experiment among many others that could have occurred. These opposing stances have important practical implications, notably the need for Bayesians to formulate their reasonable convictions or available information prior to seeing the result of an experiment, then updating this information by incorporating the data.

## B.1 Bayes formula and likelihood principle

The Bayesian paradigm rests on the following formula:

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{B.1}$$

$$\propto p(x|\theta)p(\theta) \tag{B.2}$$

where

- $x$ is the data
- $p(\theta|x)$ is the posterior distribution of the parameters
- $p(x|\theta)$ is the likelihood of the data given the parameters
- $p(\theta)$ is the prior distribution of the parameters
- $p(x|\theta)p(\theta)$ forms the unnormalised posterior density[Gelman et al., 2014]
- and

$$p(x) = \int p(\theta)p(x|\theta)d\theta \tag{B.3}$$

$p(x|\theta)$ is the familiar likelihood used in the frequentist paradigm and it is complemented by two other objects, the prior distribution which expresses the information available on the parameters without knowing the data, and the posterior distribution which combines the a priori information with evidence from the data into a density distribution for the parameters. Depending on what is known about the value of the parameters beforehand, the prior can be informative, vaguely informative or non-informative. $p(x)$ is often not considered as it does not depend on $\theta$ and will not influence the posterior distribution of the parameters which solely depend on the likelihood and the prior. This dependence entails that inference for different generating processes (e.g. binomial and negative binomial), whose likelihood have the same dependence on the parameters, will be the same[1]. This is not the case in the frequentist paradigm, notably for hypothesis testing.

## B.2 Exchangeability

Another feature of the Bayesian framework is that the assumption of independently identically distributed data used in frequentist statistics must be replaced by an exchangeability assumption. This can be understood on a simple lognormal SSD example: if $N$ species are randomly sampled from a community and if the log of the CECs in that community are assumed to follow a normal distribution of parameters $\mu$ and $\sigma$ (thus they are identically distributed), and if $\mu$ and $\sigma$ are given prior distributions, then the CECs of the sampled species are not independent.
The joint probability of observing two CECs $x$ and $y$ is:

$$p(x,y) = \int_{\mu,\sigma} p(x,y|\mu,\sigma)p(\mu,\sigma)d\mu d\sigma \tag{B.4}$$

whereas the product of the probabilities of observing $x$ and $y$ is:

$$p(x)p(y) = \int_{\mu,\sigma} p(x|\mu,\sigma)p(\mu,\sigma)d\mu d\sigma \int_{\mu',\sigma'} p(y|\mu',\sigma')p(\mu',\sigma')d\mu' d\sigma' \tag{B.5}$$

Thus, the assumption of random sampling from a normal distribution does not imply that the data is independently distributed in a Bayesian framework.
However, the assumption of random sampling from a community implies a property of exchangeability: if species are sampled randomly, then the probability of observing $x$ then $y$ is the same as that of observing $y$ then $x$, or stated differently $p(x,y) = p(y,x)$. A non random sampling would consist for example in sampling one sensitive species, then a tolerant one. Using this non random sampling, we cannot expect the probability to find a small then a large CEC to be equal to that of finding a large then a small CEC, then there is no exchangeability.
In his lecture notes, Jordan[2] gives an intuitive description of the consequences of the exchangeability assumption (more precisely of infinite exchangeability) implied through

[1] This is called the **likelihood** principle[Robert, 2007]

[2] http://www.cs.berkeley.edu/ jordan/courses/260-spring10/lectures/lecture1.pdf, a similar formulation is given in [Bernardo, 1996]

De Finetti's representation theorem[Bernardo and Smith, 2000]:

- There must exist a parameter
- There must exist a likelihood
- There must exist a prior
- Those exist so that the data is independent conditionally on the parameter

These properties are often invoked to justify the Bayesian paradigm[Bernardo, 1996]. In the particular case where the data is assumed to be exchangeable and to be sampled from a normal, then the data is independently distributed conditional on parameters $\mu$ and $\sigma$.
To come back to the simple lognormal SSD example, species exchangeability implies that the species are independently and identically distributed conditionally on parameters $\mu$ and $\sigma$. It also implies that the information available on $\mu$ and $\sigma$ can be described by their posterior density (Equation B.2).

## B.3 Numerical methods

Many numerical methods have been developed to compute the posterior distribution, relying heavily on Gibbs sampling to generate random numbers from complex joint distributions. Tools to perform these computations include JAGS[Plummer, 2013], STAN[Stan Development Team, 2014] and Winbugs[Lunn et al., 2000]. The performance of these tools and the unified framework for computing the posterior distributions make Bayesian inference a very flexible tool for modelling, able to cope with complex model structures and notably hierarchical structures very well.

## B.4 Prediction

Prediction in the Bayesian framework is easy to obtain by using Equation B.1 to define a posterior predictive distribution $p(\tilde{x}|x)$[Gelman et al., 2014]:

$$p(\tilde{x}|x) = \int p(\tilde{x}|\theta)p(\theta|x)d\theta \tag{B.6}$$

This makes the Bayesian paradigm very useful for risk assessment as it provides a straightforward method to compute the distribution of any quantity of interest, as the $HC_5$ or any other summary.

## B.5 Credible intervals and confidence intervals

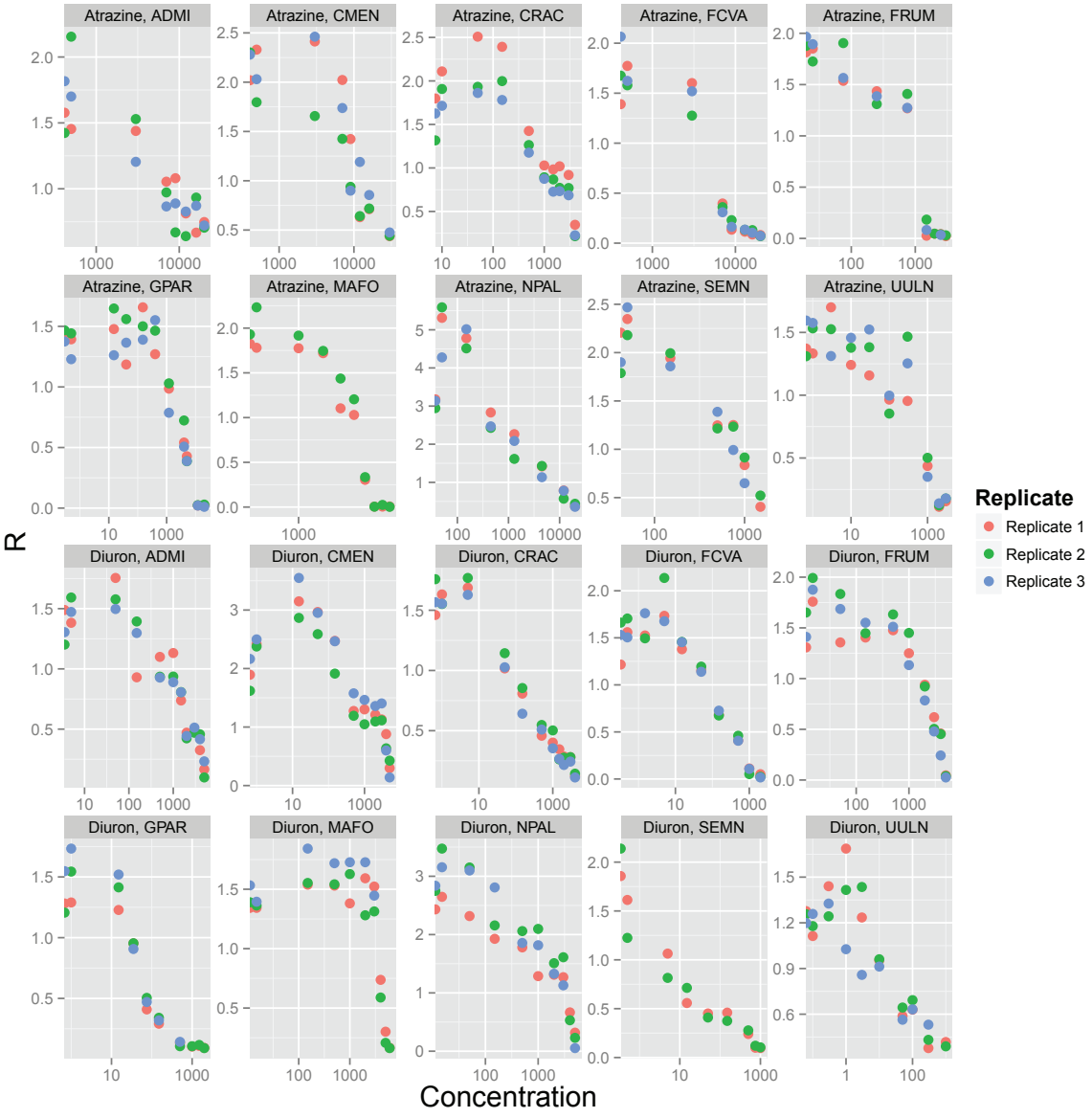
One implication of adopting a Bayesian framework is that the uncertainty on a parameter is expressed as a credible interval instead of a confidence interval in the frequentist framework. The 95% credible interval delimits a region where the parameters should

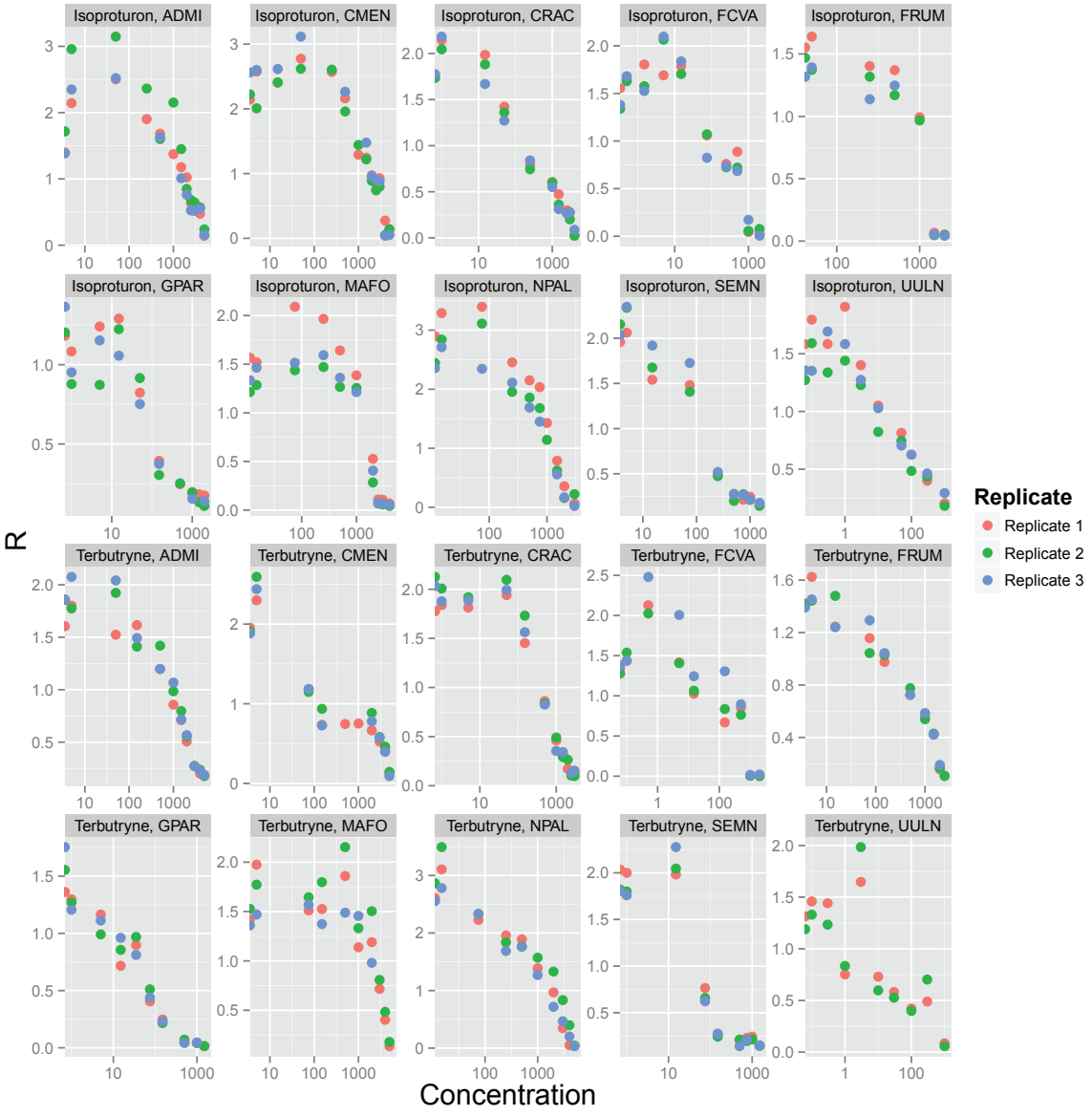
lie with 95% probability, whereas the confidence interval delimits a region which has 95% probability of containing the true value of the parameter. As such, in the context of repeated experiments the confidence interval focuses on giving a interval containing the true value 95% of the time, but can give absurd results 5% of the time, while the credible interval can fail to contain the true value most of the time if the value was really unexpected (in the sense that the available prior information is that this value should be very unlikely). In the context of fixed data (no repetition), the confidence interval may be stuck in giving an absurd interval, while the credible interval will give the region where the parameter is probably located and this region will always be reasonable provided the prior is reasonable as well.

Appendix C

# Appendix for the diatom data

## C.1 Raw data for all species and herbicides


Atrazine, ADMI
Atrazine, CMEN
Atrazine, CRAC
Atrazine, FCVA
Atrazine, FRUM
Atrazine, GPAR
Atrazine, MAFO
Atrazine, NPAL
Atrazine, SEMN
Atrazine, UULN
Diuron, ADMI
Diuron, CMEN
Diuron, CRAC
Diuron, FCVA
Diuron, FRUM
Diuron, GPAR
Diuron, MAFO
Diuron, NPAL
Diuron, SEMN
Diuron, UULN
R
Concentration
Replicate
Replicate 1
Replicate 2
Replicate 3


Isoproturon, ADMI
Isoproturon, CMEN
Isoproturon, CRAC
Isoproturon, FCVA
Isoproturon, FRUM
Isoproturon, GPAR
Isoproturon, MAFO
Isoproturon, NPAL
Isoproturon, SEMN
Isoproturon, UULN
Terbutryne, ADMI
Terbutryne, CMEN
Terbutryne, CRAC
Terbutryne, FCVA
Terbutryne, FRUM
Terbutryne, GPAR
Terbutryne, MAFO
Terbutryne, NPAL
Terbutryne, SEMN
Terbutryne, UULN
R
Concentration
Replicate
Replicate 1
Replicate 2
Replicate 3

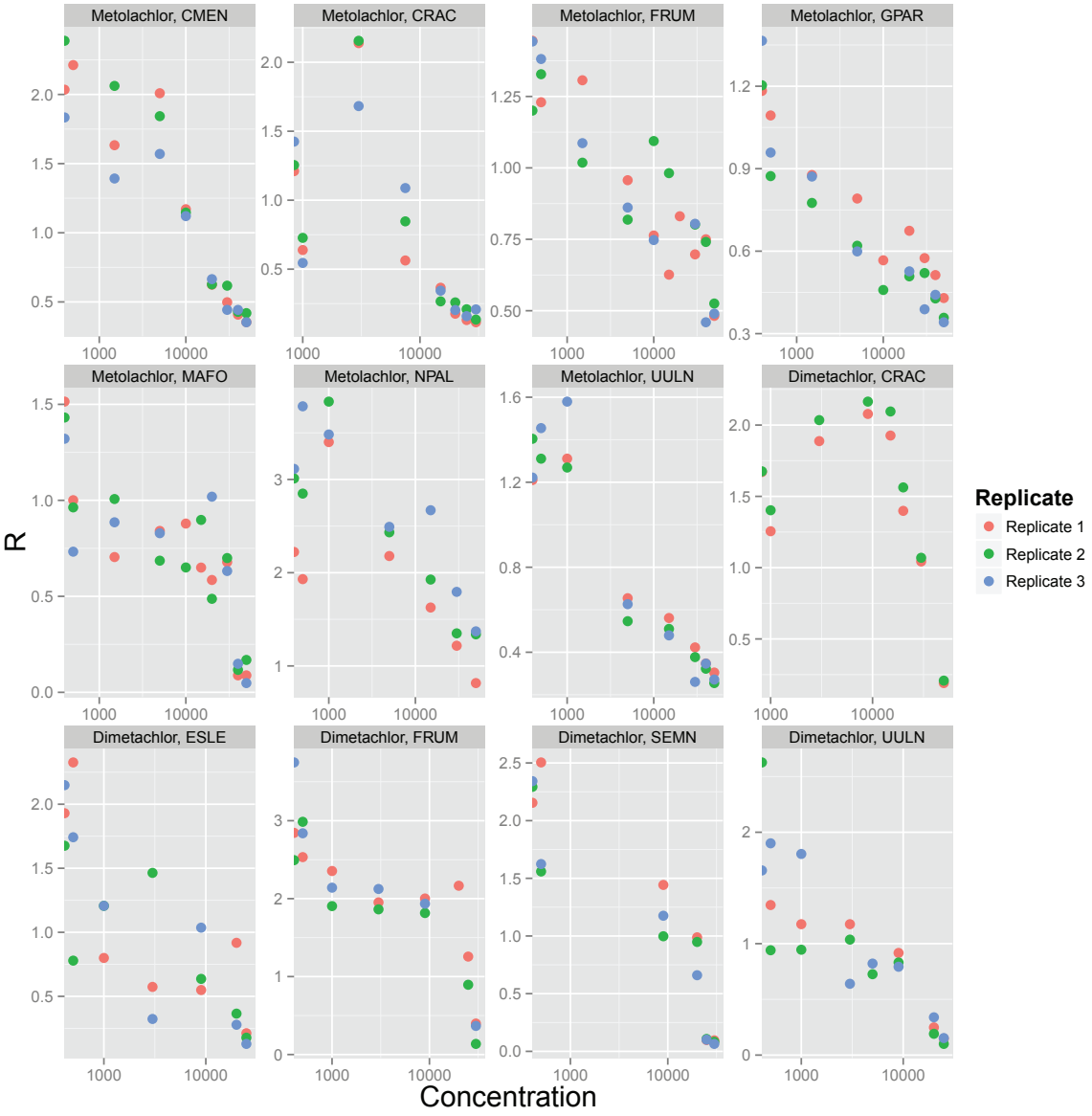
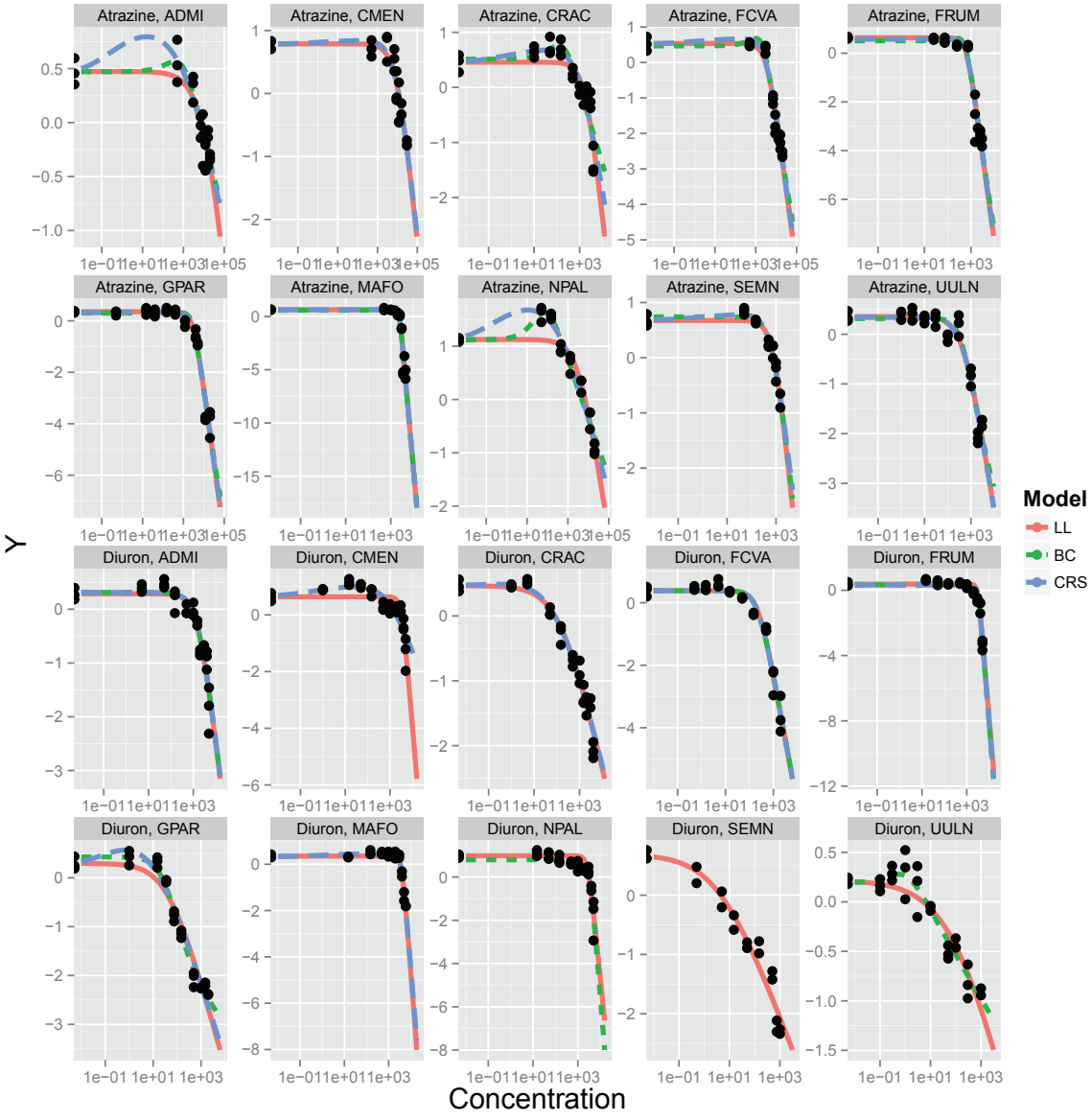
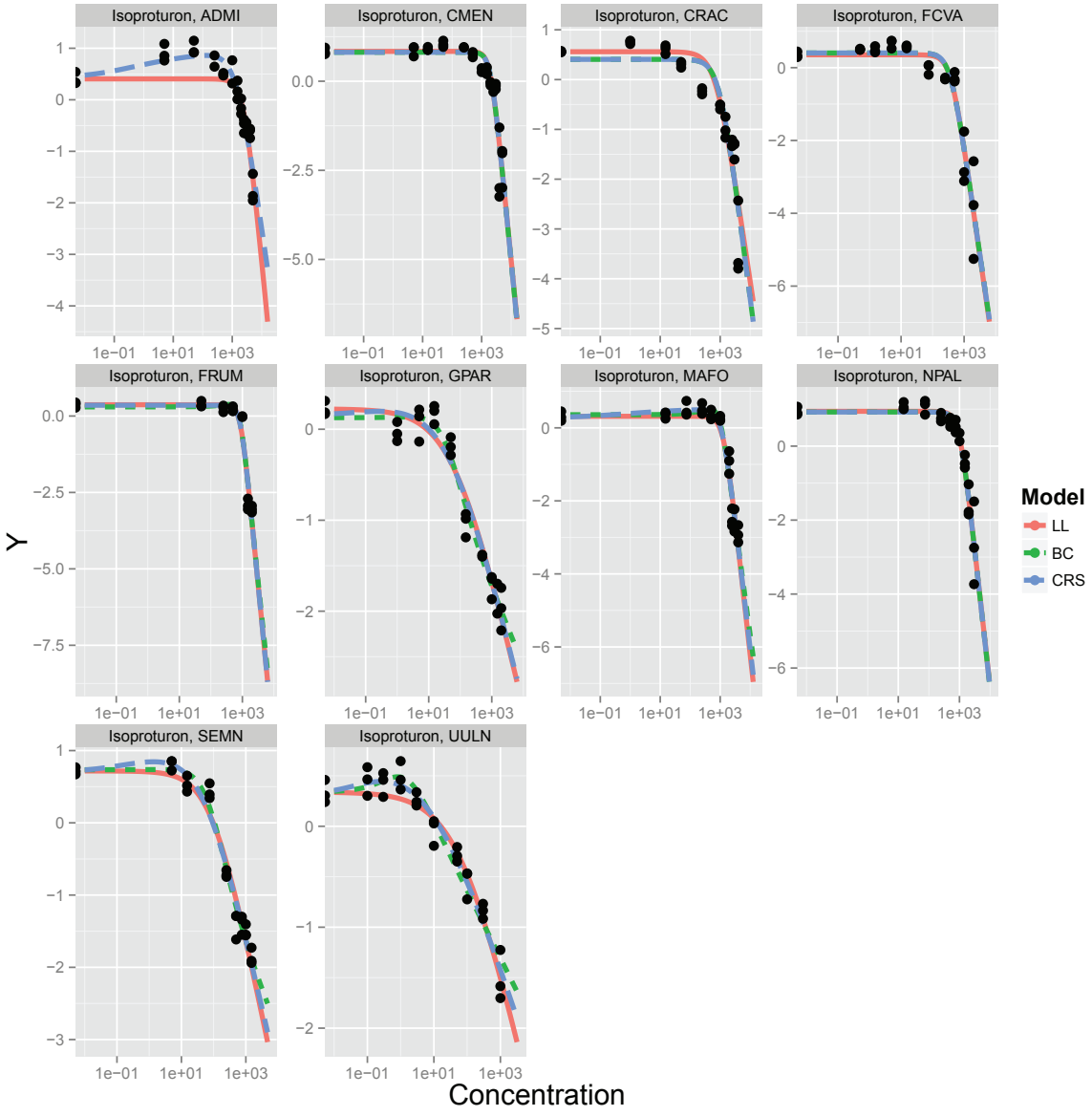



Figure C.1: Original data for all species and herbicides. $R$ is plotted against the concentration, with one colour per replicate. The concentration is in log-scale and the control measurements (zero concentration) are drawn on the left border of the plot.

## C.2 Fit of the three models for all species and herbicides


Atrazine, ADMI
Atrazine, CMEN
Atrazine, CRAC
Atrazine, FCVA
Atrazine, FRUM
Atrazine, GPAR
Atrazine, MAFO
Atrazine, NPAL
Atrazine, SEMN
Atrazine, UULN
Diuron, ADMI
Diuron, CMEN
Diuron, CRAC
Diuron, FCVA
Diuron, FRUM
Diuron, GPAR
Diuron, MAFO
Diuron, NPAL
Diuron, SEMN
Diuron, UULN
Y
Concentration
Model
LL
BC
CRS


Isoproturon, ADMI
Isoproturon, CMEN
Isoproturon, CRAC
Isoproturon, FCVA
Isoproturon, FRUM
Isoproturon, GPAR
Isoproturon, MAFO
Isoproturon, NPAL
Isoproturon, SEMN
Isoproturon, UULN
Y
Concentration
1e−01
1e+01
1e+03
Model
LL
BC
CRS

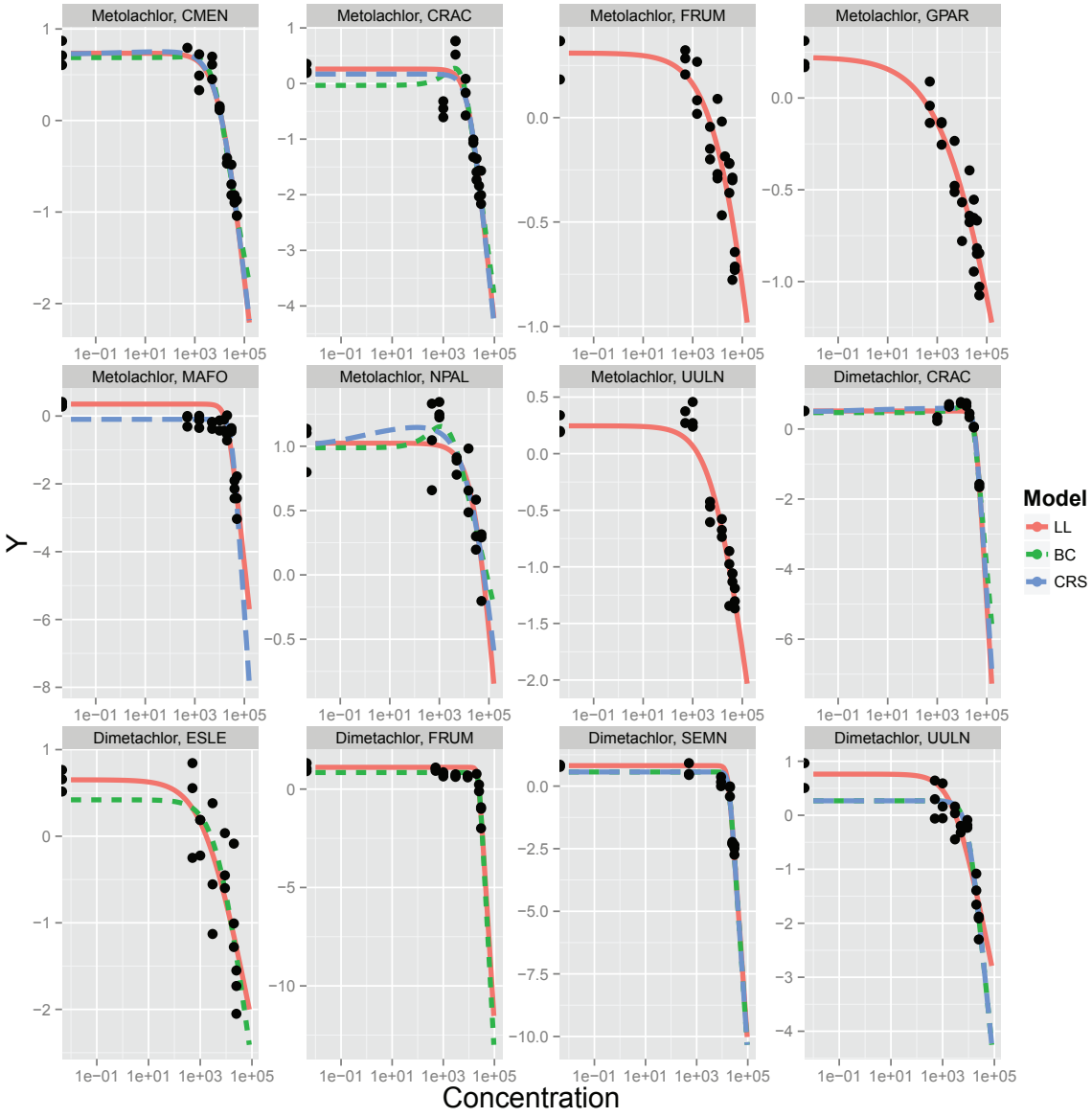



Figure C.2: Comparison of the fit of the three models on the log-transformed data. LL stands for log-logistic, BC for Brain-Cousens and CRS for Cedergreen-Ritz-Streibig.

## C.3 Akaike Information Criterion for the three concentration-effect models

| Pesticide | Species | AIC LL | AIC BC | AIC CRS | AIC CRS-AIC LL | AIC BC-AIC LL |
|---|---|---|---|---|---|---|
| Atrazine | NPAL | 11 | -5 | -9 | -21 | -16 |
| Terbutryne | NPAL | 42 | 40 | 40 | -2 | -2 |
| Diuron | NPAL | 33 | 33 | NA | NA | 0 |
| Isoproturon | NPAL | 29 | 33 | 33 | 4 | 4 |
| Metolachlor | NPAL | 0 | 2 | 2 | 2 | 2 |
| Atrazine | ADMI | -15 | -17 | -16 | -1 | -2 |
| Terbutryne | ADMI | -49 | -45 | -45 | 4 | 4 |
| Diuron | ADMI | 6 | 9 | 9 | 4 | 4 |
| Isoproturon | ADMI | 26 | NA | 9 | -17 | NA |
| Atrazine | FCVA | 7 | 2 | 9 | 3 | -5 |
| Terbutryne | FCVA | 69 | NA | NA | NA | NA |
| Diuron | FCVA | 15 | 19 | 19 | 4 | 4 |
| Isoproturon | FCVA | 48 | 52 | 52 | 4 | 4 |
| Atrazine | CMEN | -1 | NA | 3 | 4 | NA |
| Terbutryne | CMEN | 36 | NA | NA | NA | NA |
| Diuron | CMEN | 29 | NA | 32 | 3 | NA |
| Isoproturon | CMEN | 51 | 55 | 55 | 4 | 4 |
| Metolachlor | CMEN | -22 | -22 | -18 | 4 | 0 |
| Atrazine | MAFO | 52 | 56 | 56 | 4 | 4 |
| Terbutryne | MAFO | -14 | -14 | -14 | 0 | 0 |
| Diuron | MAFO | -17 | NA | -22 | -5 | NA |
| Isoproturon | MAFO | 25 | 24 | 26 | 0 | -1 |
| Metolachlor | MAFO | 44 | NA | 32 | -12 | NA |
| Atrazine | FRUM | 37 | 39 | 41 | 4 | 2 |
| Terbutryne | FRUM | -5 | -7 | -7 | -2 | -2 |
| Diuron | FRUM | 26 | 29 | 29 | 4 | 4 |
| Isoproturon | FRUM | 25 | 29 | 29 | 4 | 4 |
| Metolachlor | FRUM | -19 | NA | NA | NA | NA |
| Atrazine | GPAR | 25 | 28 | 29 | 4 | 2 |
| Terbutryne | GPAR | 4 | 3 | 3 | -1 | -1 |
| Diuron | GPAR | 10 | -16 | 4 | -6 | -26 |
| Isoproturon | GPAR | -5 | -14 | -2 | 3 | -8 |
| Metolachlor | GPAR | -28 | NA | NA | NA | NA |
| Atrazine | UULN | 6 | 9 | 10 | 4 | 2 |
| Terbutryne | UULN | 28 | 35 | 32 | 4 | 7 |
| Diuron | UULN | -17 | -26 | NA | NA | -10 |
| Isoproturon | UULN | -22 | -19 | -27 | -5 | 3 |
| Metolachlor | UULN | -9 | NA | NA | NA | NA |
| Atrazine | CRAC | 13 | 16 | 12 | -1 | 3 |
| Terbutryne | CRAC | -27 | -23 | -23 | 4 | 4 |
| Diuron | CRAC | -17 | NA | -14 | 4 | NA |
| Isoproturon | CRAC | 46 | 49 | 49 | 3 | 3 |
| Metolachlor | CRAC | 26 | 26 | 30 | 4 | 0 |
| Atrazine | SEMN | -20 | -18 | -20 | 0 | 2 |
| Terbutryne | SEMN | 23 | 10 | NA | NA | -13 |
| Diuron | SEMN | 5 | NA | NA | NA | NA |
| Isoproturon | SEMN | -4 | -13 | -4 | 0 | -9 |
| Dimetachlor | SEMN | 26 | 27 | 27 | 1 | 1 |
| Dimetachlor | CRAC | -8 | -11 | -7 | 2 | -2 |
| Dimetachlor | UULN | 24 | 21 | 21 | -4 | -4 |
| Dimetachlor | ESLE | 35 | 39 | NA | NA | 5 |
| Dimetachlor | FRUM | 21 | 13 | NA | NA | -8 |

Table C.1: AIC values for each of the three concentration-effect models for the diatom dataset. The last two columns provide the $\Delta$AIC for model comparison.

Appendix D

# JAGS code for the diatom log-logistic hierarchical model

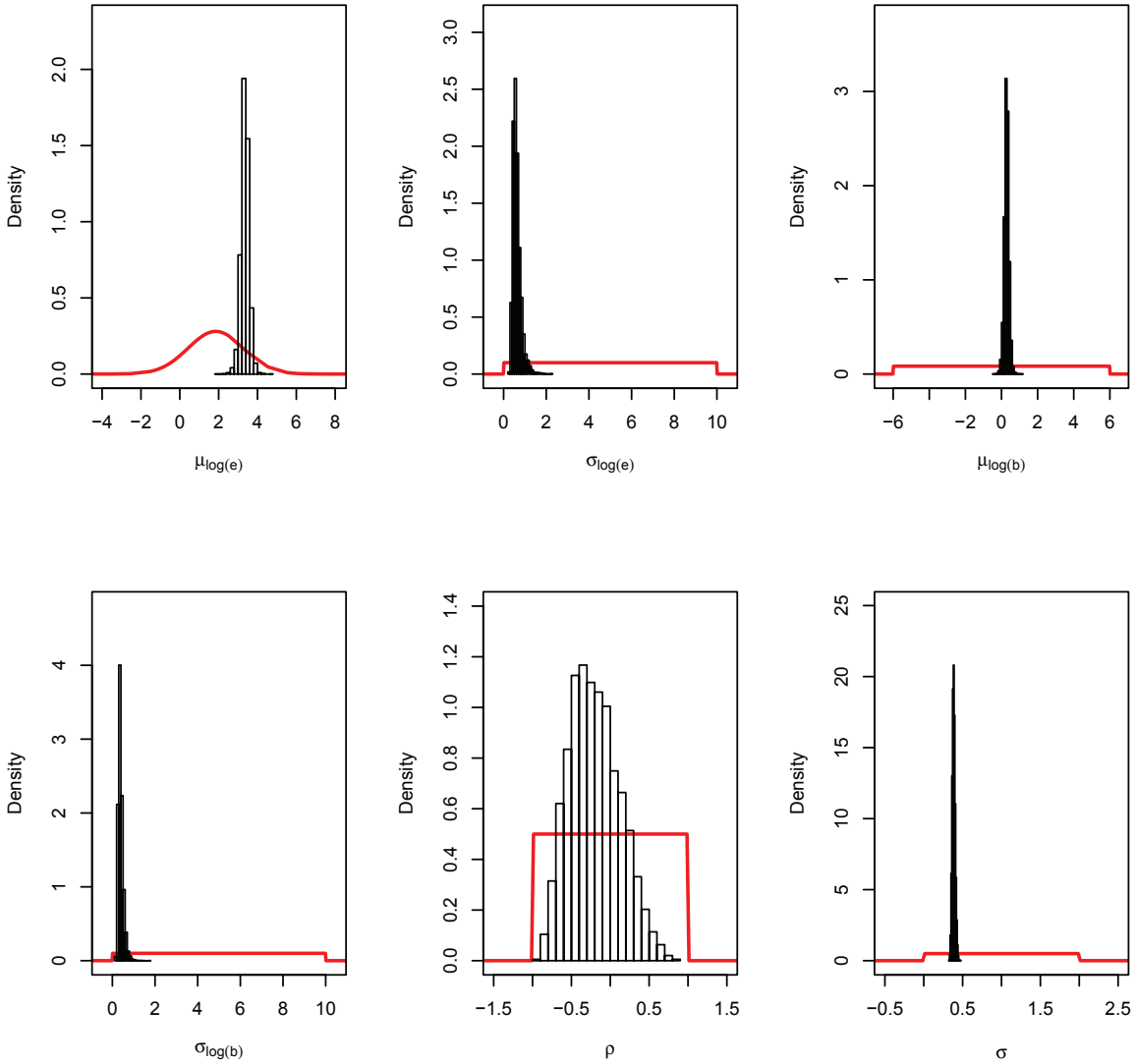
```
model
{
  # Data and error model

  for (i in 1:ndat)
  {
    R[i] <- log(d[species[i]]  /
                  (1. + (concentration[i] / e[species[i]]) ^ b[species[i]]))
    log_fluo[i] ~ dnorm(R[i] , tau)
  }

  # One set of parameters per species

  for (j in 1:nspecies){

    e[j] <- 10 ^ le[j]
    b[j] <- 10 ^ lb[j]
    lb[j] <- B[j , 1]
    le[j] <- B[j , 2]

    B.hat[j , 1] <- lb.mu
    B.hat[j , 2] <- le.mu

  B[j , 1:2] ~ dmnorm(B.hat[j , ] , Tau.B[ , ])
  }
```

```
  # Parameter transformation

  tau  <-  1 / sigma ^ 2

  Tau.B[1:2 , 1:2] <- inverse(VarCovar.B[ , ])
  VarCovar.B[1 , 1] <- lb.sigma ^ 2
  VarCovar.B[2 , 2] <- le.sigma ^ 2
  VarCovar.B[1 , 2] <- rho * lb.sigma * le.sigma
  VarCovar.B[2 , 1] <- VarCovar.B[1 , 2]

# Prior specification

  le.mu ~ dnorm(lC.mu , 1. / lC.sigma ^ 2)
  le.sigma ~ dunif(0 , 10)
  lb.mu ~ dunif(-6. , 6.)
  lb.sigma ~ dunif(0 , 10)
  rho ~ dunif(-1. , 1.)
  sigma  ~  dunif(0. , 2.)

}
```

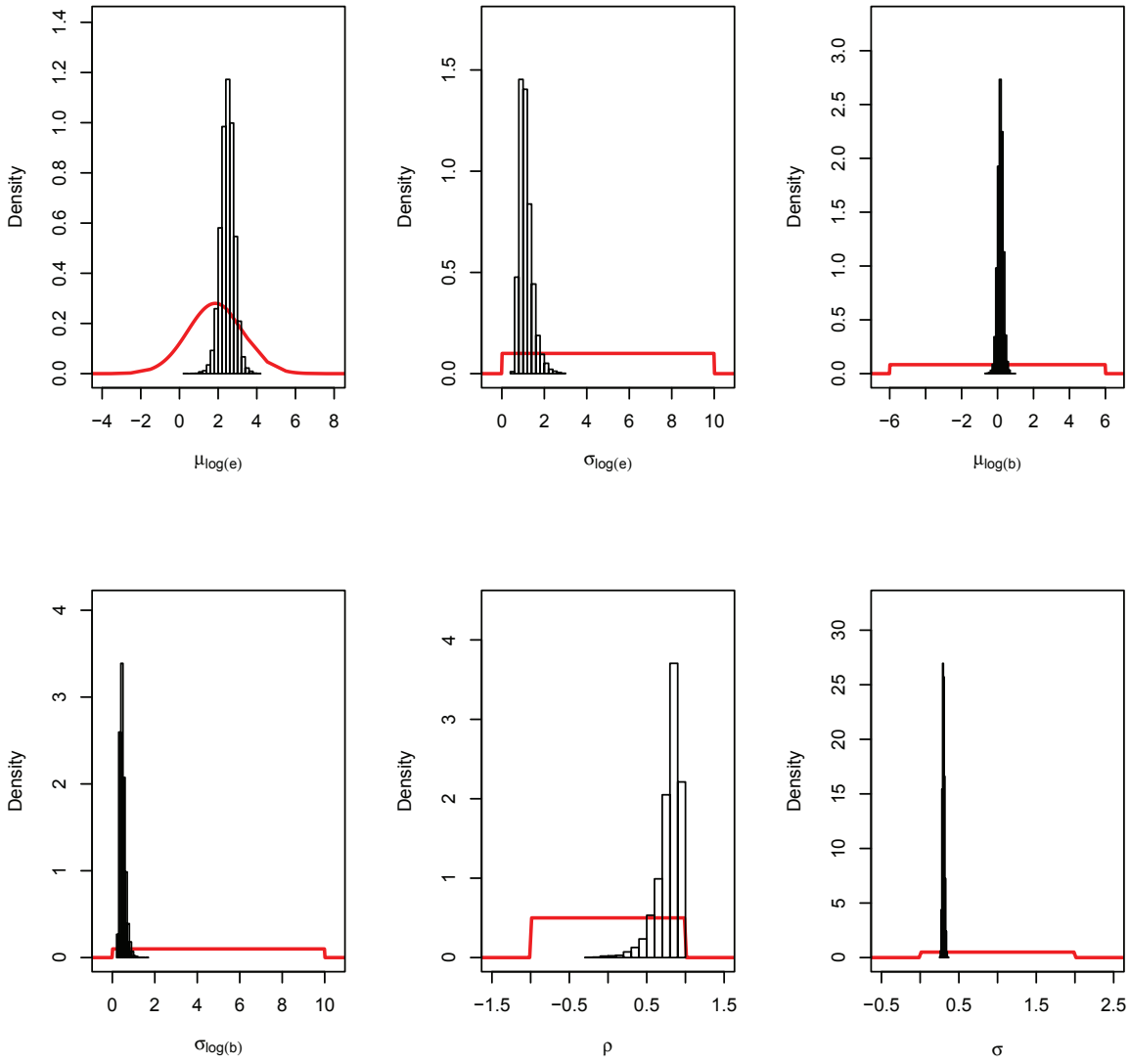
# Appendix E

# Prior-posterior plots for all herbicides

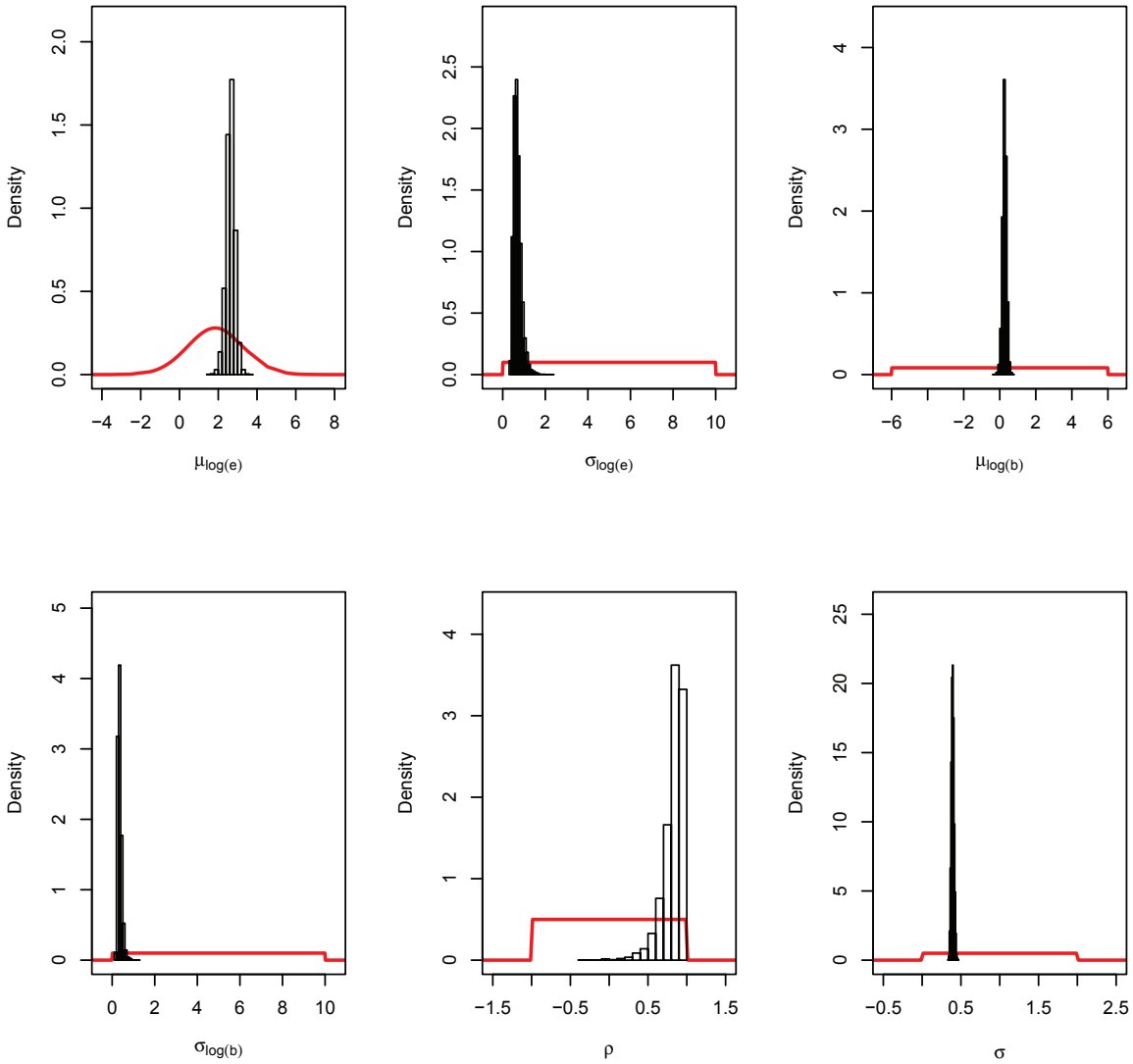
## E.1 Atrazine



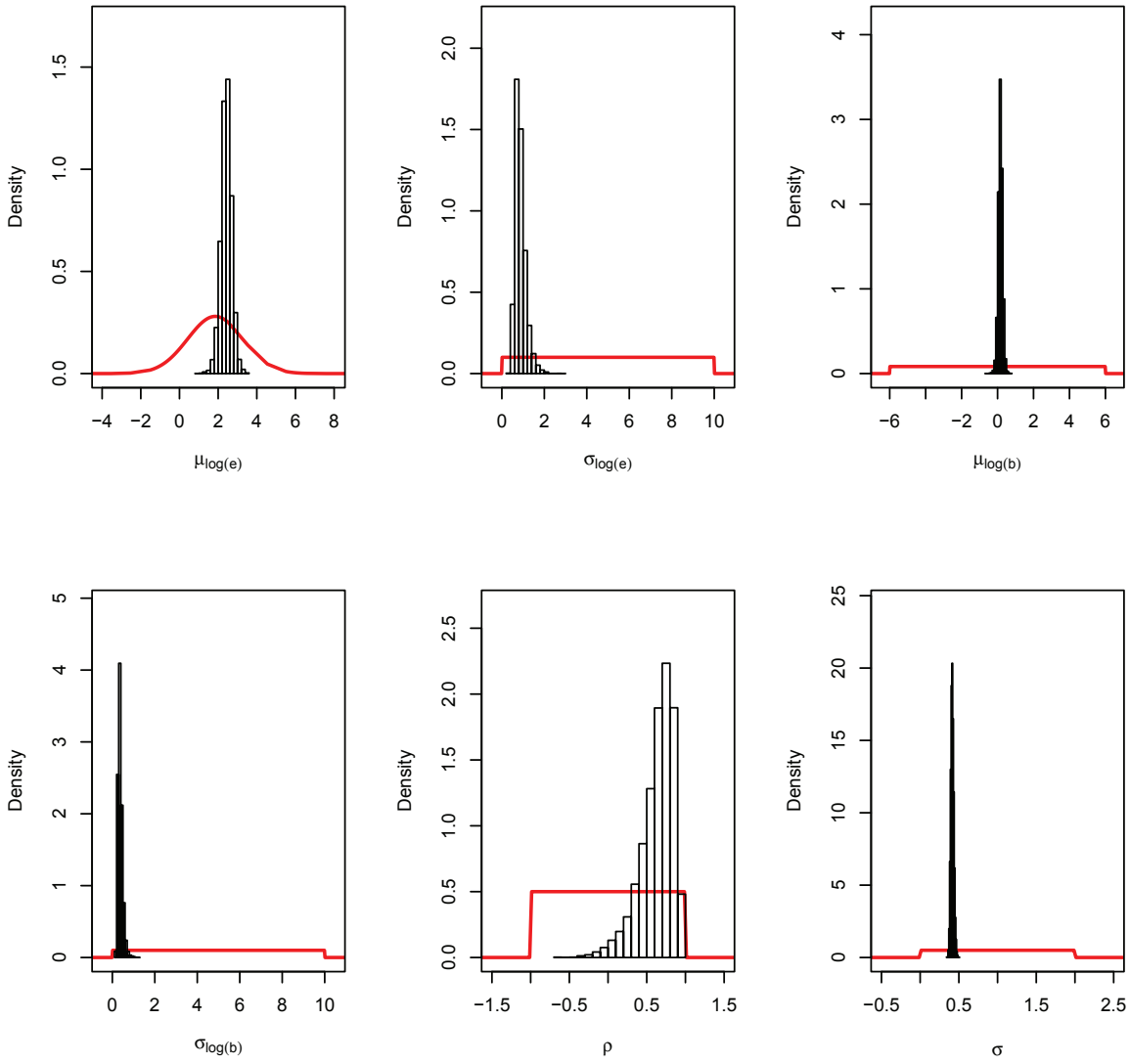
## E.2 Diuron



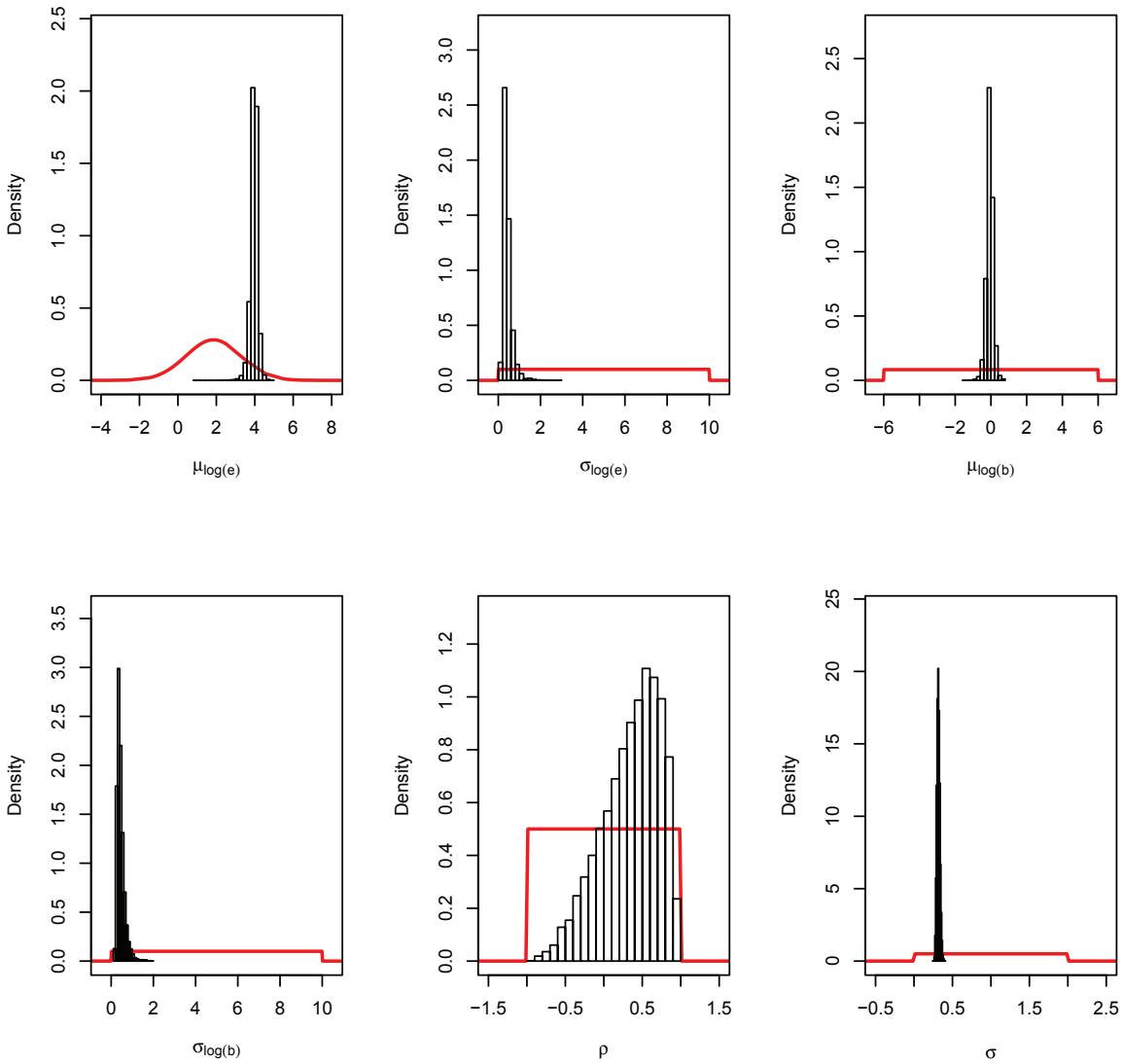
## E.3 Isoproturon



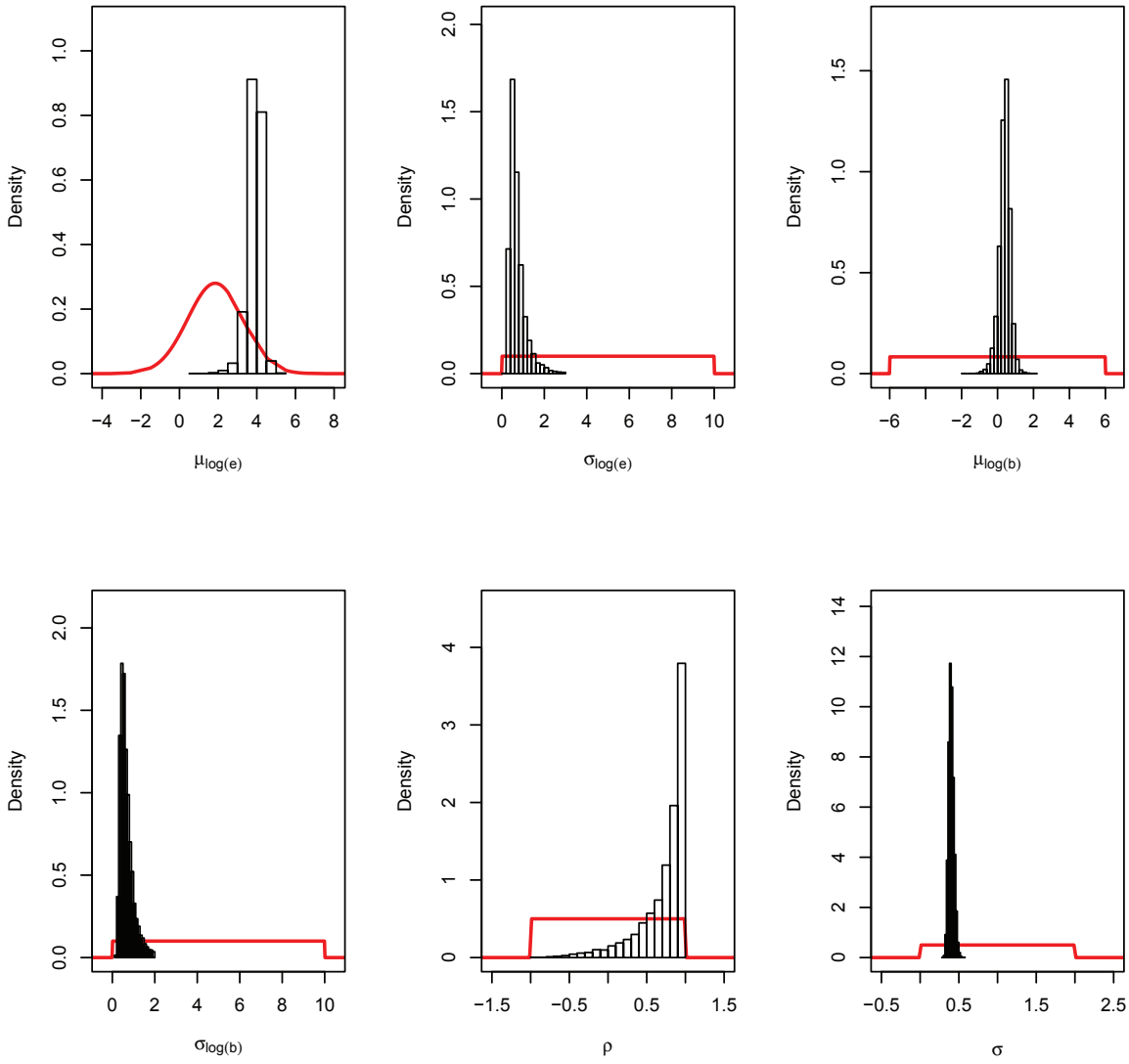
## E.4 Terbutryne



## E.5 Metolachlor



## E.6 Dimetachlor



Appendix F

# Multinomial and conditional binomial formulations of the error model

We wish to prove that the multinomial and the conditional binomial formulations of the error model are equivalent when there are no lost to follow-up organisms. To do so, we compute the likelihood of the data given the parameters using a multinomial error model and a conditional error model.

## F.1 Multinomial formulation of the model

Let $\boldsymbol{\theta}$ be the parameters of the TKTD model for one species, $T$ the time at which the death of an organism of that species occurs and $S_k(\boldsymbol{\theta}) = P(T > t_k)$ the probability for that organism to survive until time $t_k$ given $\boldsymbol{\theta}$.
Let $N_0$ be the initial number of organisms, $l$ the number of measurement dates, $N_k$ the number of organisms alive at time $t_k$ and $m_k$ the multinomial probability for an organism to die between $t_{k-1}$ and $t_k$.
$\forall k \in \mathbb{N}_l$;

$$\begin{aligned} m_k =& P(t_{k-1} < T < t_k) && \text{(F.1)} \\ =& P(T < t_k) - P(T < t_{k-1}) && \text{(F.2)} \\ =& 1 - P(T > t_k) - (1 - P(T > t_{k-1})) && \text{(F.3)} \\ =& S_{k-1}(\boldsymbol{\theta}) - S_k(\boldsymbol{\theta}) && \text{(F.4)} \end{aligned}$$

and the probability to die after the last measurement is:

$$m_{l+1} = S_l(\boldsymbol{\theta}) \tag{F.5}$$

The number of organisms dying between $t_{k-1}$ and $t_k$ is $\Delta_k = N_{k-1} - N_k$, and the number of organisms dying after $t_l$ is $\Delta_{l+1} = N_l$, so the multinomial distribution for the values $\mathbf{N} = (N_1 \dots N_l)$ given $N_0$ and $\boldsymbol{\theta}$ is:

$$f(\mathbf{N}|N_0,\boldsymbol{\theta}) = N_0! \prod_{k=1}^{l+1} \frac{m_k^{\Delta_k}}{\Delta_k!} \tag{F.6}$$

$$= \frac{N_0! S_l^{N_l}}{N_l!} \prod_{k=1}^{l} \frac{(S_{k-1}-S_k)^{N_{k-1}-N_k}}{(N_{k-1}-N_k)!} \tag{F.7}$$

where the dependence of $S_k$ on $\boldsymbol{\theta}$ was omitted for readability.
In this formulation, there are two necessary constraints associated to the use of the multinomial distribution:

$$\sum_{k=0}^{l+1} \Delta_k = N_0 \tag{F.8}$$

and

$$\sum_{k=0}^{l+1} m_k = 1 \tag{F.9}$$

## F.2 Conditional binomial formulation of the model

We use the same notations as before and define $p_k$ the probability to survive until $t_k$ for an organism alive at $t_{k-1}$:

$$p_k = \frac{S_k(\boldsymbol{\theta})}{S_{k-1}(\boldsymbol{\theta})} \tag{F.10}$$

Note that $p_k$ was noted $S\left(t_k|t_{k-1}\right)$ in the main text, but we use this shorter notation here. In the absence of lost to follow-up organisms, the conditional binomial relation is:

$$f(N_k|N_{k-1},\boldsymbol{\theta}) = \frac{N_{k-1}!}{N_k!(N_{k-1}-N_k)!} p_k^{N_k} (1-p_k)^{N_{k-1}-N_k} \tag{F.11}$$

Notice that in this formulation, $N_{k-1}$ is both a variable predicted by the model and a covariable for $N_k$. The full conditional binomial distribution is given by:

$$f(\mathbf{N}|N_0,\boldsymbol{\theta}) = \prod_{k=1}^{l} \frac{N_{k-1}!}{N_k!(N_{k-1}-N_k)!} p_k^{N_k}(1-p_k)^{N_{k-1}-N_k} \tag{F.12}$$

$$= \frac{N_0}{N_l!} \prod_{k=1}^{l} \frac{1}{(N_{k-1}-N_k)!} \prod_{k=1}^{l} \left(\frac{S_k}{S_{k-1}}\right)^{N_k} (1-\frac{S_k}{S_{k-1}})^{N_{k-1}-N_k} \tag{F.13}$$

$$= \frac{N_0}{N_l!} \prod_{k=1}^{l} \frac{1}{(N_{k-1}-N_k)!} \prod_{k=1}^{l} \frac{S_k^{N_k}}{S_{k-1}^{N_{k-1}}} (S_{k-1}-S_k)^{N_{k-1}-N_k} \tag{F.14}$$

$$= \frac{N_0 S_l^{N_l}}{N_l! S_0^{N_0}} \prod_{k=1}^{l} \frac{(S_{k-1}-S_k)^{N_{k-1}-N_k}}{(N_{k-1}-N_k)!} \tag{F.15}$$

$$= \frac{N_0 S_l^{N_l}}{N_l!} \prod_{k=1}^{l} \frac{(S_{k-1}-S_k)^{N_{k-1}-N_k}}{(N_{k-1}-N_k)!} \tag{F.16}$$

Equation F.7 and Equation F.16 are identical, showing that the two formulations are equivalent. In the conditional binomial formulation, it is possible to account for the lost to follow-up organisms (which were eaten, or which emerged from the tank) by replacing $N_{k-1}$ by $N_k^p$ in Equation F.11, where $N_k^p$ is the number of organisms alive at the previous time minus the number of lost to follow-up organisms. Replacing the multinomial error model in the presence of lost to follow-up organisms makes the assumption that the mechanism by which they are lost is not related to toxicity in that it does not influence the estimation of the parameters.

Appendix G

# Limit cases for the General Unified Threshold model for Survival

## G.1 General Unified Threshold model for Survival - Stochastic Death equations with damage:

$$\dot{C}_i = k_i C_w - k_e C_i \tag{G.1}$$
$$\dot{D} = k_a C_i - k_r D \tag{G.2}$$
$$\tag{G.3}$$

This model has three compartments, $C_w$, $C_i$ and $D$.
In the following derivation, we assume that the external concentration $C_w$ is constant. Indeed, here the aim is to find a closed form for the survival probability to estimate the parameters.
The hazard rate is supposed to be linearly dependant on damage :

$$h_z = m^0 + k_D^s \left(D - D_0\right)_+ \tag{G.4}$$

As the only the external concentration is known, we want to use it as the dose-metric.

## G.2 Steady-state approximation

[Jager et al., 2011] explain that the slowest compartment will dominate the dynamic and that only the slowest rate constant will be identifiable.
Let's assume that the internal concentration is the fastest evolving compartment. We can then assume that with respect to the evolution of the damage compartment, the internal concentration is always in steady state, a typical approximation in chemical kinetics.
This assumption translates to:

$$\dot{C}_i = 0 \tag{G.5}$$

$$\implies C_i = \frac{k_i}{k_e} C_w \tag{G.6}$$

Substituting Equation G.6 in Equation G.3 yields :

$$\dot{D} = k_a \frac{k_i}{k_e} C_w - k_r D \tag{G.7}$$

With this assumption, the Toxico-Kinetic and Toxico-Dynamic processes reduce to the evolution of the damage compartment.
If instead we assume that the damage compartment evolves fastest, we get:

$$\dot{D} = 0 \tag{G.8}$$

$$\implies D = \frac{k_a}{k_r} C_i \tag{G.9}$$

With this assumption, all the Toxico-Kinetic and Toxico-Dynamic processes reduce to the evolution of the internal concentration compartment.

## G.3 Scaling of the dose-metric to solve the identifiability problem

Two of these compartments, $C_i$ and $D$, are not measured in the most common type of survival data. [Jager et al., 2011] explain that unobserved compartments prevent the estimation of parameters $k_i$ and $k_a$, but that these can be rescaled out by modelling a scaled internal concentration and a scaled damage.
Let's write Equation G.2 with a scaled internal concentration :

$$\dot{C}_i^* = C_w - k_e C_i^* \tag{G.10}$$

where

$$C_i^* = \frac{C_i}{k_i} \tag{G.11}$$

Under the assumption that the damage compartment evolves fastest (Equation G.9), the hazard rate (Equation G.4) can be written:

$$h_z = m^0 + k_D^s \left( \frac{k_a k_i}{k_r} C_i^* - D_0 \right)_+ \tag{G.12}$$

$$= m^0 + k^s \left( C_i^* - NEC \right)_+ \tag{G.13}$$

where $k^s = \frac{k_a k_i}{k_r} k_D^s$ and $NEC = \frac{k_r}{k_a k_i} D_0$.
As $S(t) = e^{-\int_0^t h_z(u)du}$, to compute the survival probability for a given $C_w$ we need to integrate Equation G.10 so we need to know the scaled internal concentration at one date. The most likely assumption in the absence of information on the internal concentration is $C_i^*(0) = 0$.
With this assumption:

$$C_i^*(t) = C_w(1 - e^{-k_e t}) \tag{G.14}$$

and Equation G.13 becomes:

$$h_z = m^0 + k^s \left(C_w(1 - e^{-k_e t}) - NEC\right)_+ \tag{G.15}$$

Under the assumption that the internal concentration compartment evolves fastest (Equation G.7), Equation G.3 can be written:

$$\dot{D^*} = C_w - k_r D^* \tag{G.16}$$

where

$$D^* = \frac{k_e}{k_a k_i} D \tag{G.17}$$

It seems safer to assume that the initial damage is negligible ($D(0) = 0 \implies D^*(0) = 0$) which yields:

$$D^*(t) = C_w(1 - e^{-k_r t}) \tag{G.18}$$

Replacing Equation G.18 and Equation G.17 in Equation G.4 yields:

$$h_z = m^0 + k_D^s \left(\frac{k_a k_i}{k_e} C_w(1 - e^{-k_r t}) - D_0\right)_+ \tag{G.19}$$

$$= m^0 + k^s \left(C_w(1 - e^{-k_r t}) - NEC\right)_+ \tag{G.20}$$

where $k^s = \frac{k_a k_i}{k_e} k_D^s$ and $NEC = \frac{k_e}{k_a k_i} D_0$.
Equation G.13 and Equation G.20 are identical and use $C_w$ as the dose metric. Only the expression of the parameters $k^s$, $NEC$ and the rate constant in the exponential term change. Both imply a four-parameters model which we have found to provide a good description of survival data. The comment that only the slowest rate constant is identifiable holds true.
However, we see that the assumption of negligible initial internal concentration is needed when the damage compartment evolves fastest, whereas the assumption of negligible initial damage (conceptually easier) is needed when the internal concentration compartment evolves fastest.
In the general case, there is no good reason to make any other assumption than $D(0) = 0$,

therefore the GUTS model is not very suited when one cannot make the assumption $C_i(0) = 0$.

Appendix H

# Stan code for the time-resolved Species Sensitivity Distribution based on a hierarchical Toxico Dynamic model

```
data {
  int<lower = 0> ndat; # Number of data points
  int<lower = 0> nspecies; # Number of species
  real<lower = 0> minc; # minimum concentration
  real<lower = 0> maxc; # maximum concentration
  real<lower = 0> x[ndat]; # Concentration
  real<lower = 0> t[ndat]; # measurement date
  real<lower = 0> tprec[ndat]; # previous measurement date
  int<lower = 0> y[ndat]; # number of survivors
  int<lower = 0> Nprec[ndat]; # Initial number of species
  int<lower = 0> species[ndat]; # index of species number
}

parameters {#all the hyperparameters
  real<lower = -7,upper = 2> lks_mu;
  real<lower = -7,upper = 2> lkr_mu;
  real<lower = -7, upper = 2> lm0_mu;
  real<lower = log(minc)/log(10)-1,upper = log(maxc)/log(10)+1> lNEC_mu;
  vector<lower = 0,upper = 3>[4] tau; # prior scale
  matrix[4,nspecies] z;
  cholesky_factor_corr[4] L_Omega;#Cholesky factor
                                  #of the correlation matrix
}

transformed parameters{
```

```
  vector[nspecies] m0;
  vector[nspecies] ks;
  vector[nspecies] NEC;
  vector[nspecies] kr;
  matrix[nspecies,4] lparams; #matrix for the log of the parameters,
                              #intermediary variable used
                              #in the matrix products
  matrix[nspecies,4] Mu; #Position vector for the multivariate normal

  for (i in 1:nspecies){#initialization of the position vector
    Mu[i,1] <- lks_mu;
    Mu[i,2] <- lNEC_mu;
    Mu[i,3] <- lkr_mu;
    Mu[i,4] <- lm0_mu;
  }

  lparams <- Mu + (diag_pre_multiply(tau, L_Omega)*z)'; #algebraic relation
                                                        #implementing
                                      #lparams ~ normal_multivariate(Mu, Sigma)

  for (i in 1:nspecies){
    #translation from the matrix to the parameters for readability of the TKTD model
    #there is no vectorised version of the power function in Stan yet
    ks[i] <- pow(10.,lparams[i,1]);
    NEC[i] <- pow(10.,lparams[i,2]);
    kr[i] <- pow(10.,lparams[i,3]);
    m0[i] <- pow(10.,lparams[i,4]);
  }

}

model {
  vector<lower = 0,upper = 1>[ndat] psurv;
  vector<lower = 0>[ndat] tNEC;
  vector<lower = 0>[ndat] tref;

  to_vector(z) ~ normal(0,1);
  #uniform prior on all correlation matrices
  L_Omega ~ lkj_corr_cholesky(1);

  for (i in 1:ndat)  {
    #Effect of the control mortality parameter
    psurv[i] <- exp(-m0[species[i]] * (t[i]-tprec[i]) );
```

```
    if (x[i] > NEC[species[i]]){
      #Time for the internal concentration to reach the NEC
      tNEC[i] <- -1/kr[species[i]]*log(1-NEC[species[i]]/x[i]);

      if (t[i] > tNEC[i]){
      #The effect is counted either from the previous measurement time
      #or from the NEC
        tref[i] <- fmax(tprec[i],tNEC[i]);
        psurv[i] <- psurv[i]*
                    exp(- ks[species[i]]*
                          ( (x[i]-NEC[species[i]]) * (t[i]-tref[i]) +
                            1/kr[species[i]] *
                                  x[i] *
                                  ( exp(-kr[species[i]]*t[i]) -
                                        exp(-kr[species[i]]*tref[i])) ));
  #Add the contribution of the salinity to survival probability
      }
    }

  }

  y ~ binomial(Nprec,psurv);
}

generated quantities{
  #Computation of the traditional parameters
# of the multivariate normal distribution from the cholesky parametrization
  #This is optional, these computations can be done directly
# on the posterior distributions

  matrix[4,4] Sigma;
  real lks_sigma;
  real lNEC_sigma;
  real lkr_sigma;
  real lm0_sigma;
  real rho_lks_lNEC;
  real rho_lks_lkr;
  real rho_lks_lm0;
  real rho_lNEC_lkr;
  real rho_lNEC_lm0;
  real rho_lkr_lm0;

  #variance covariance matrix
```

```
  Sigma <- quad_form_diag(tcrossprod(L_Omega), tau);


  #scale parameters
  lks_sigma <- sqrt(Sigma[1,1]);
  lNEC_sigma <- sqrt(Sigma[2,2]);
  lkr_sigma <- sqrt(Sigma[3,3]);
  lm0_sigma <- sqrt(Sigma[4,4]);


  #Correlation parameters
  rho_lks_lNEC <- Sigma[1,2]/sqrt(Sigma[1,1]*Sigma[3,3]);
  rho_lks_lkr <- Sigma[1,3]/sqrt(Sigma[1,1]*Sigma[3,3]);
  rho_lks_lm0 <- Sigma[1,4]/sqrt(Sigma[1,1]*Sigma[4,4]);
  rho_lNEC_lkr <- Sigma[2,3]/sqrt(Sigma[2,2]*Sigma[3,3]);
  rho_lNEC_lm0 <- Sigma[2,4]/sqrt(Sigma[2,2]*Sigma[4,4]);
  rho_lkr_lm0 <- Sigma[3,4]/sqrt(Sigma[3,3]*Sigma[4,4]);
}
```

# Glossary

**Acute to Chronic Ratio** A factor to convert acute sensitivity to chronic sensitivity. 100, 167

**Akaike Information Criterion** A score based on deviance taking the number of parameters into account which allow to rank statistical models for a given dataset [Akaike, 1974]. 13, 167

**Approximate Bayesian Computation** A range of methods and algorithms to deal with models for which the likelihood is intractable, either because there is no analytical form or because it takes too long to compute. 106, 167

**Critical Effect Concentration** General name for ECx, LCx, NEC, NOEC, LOEC .... It is the concentration which is used to summarise a species tolerance and is most often a parameter of a function of the parameters of a concentration - response model. 3, 167

**Cumulative Distribution Function** Let $X \sim f$. $F(q) = P(X \leq Q)$ is the Cumulative Distribution Function of $f$. 7, 167

**Deviance Information Criterion** The DIC is a generalisation of the AIC used for model selection in the Bayesian framework. It includes the deviance and a penalisation for model complexity. 20, 167

$E_TX$ An RIVM sofware to fit an SSD and compute the $HC_5$. 167

**Effect Concentration at** 50% A type of CEC, the concentration which induces an effect of 50% compared to the control experiment. 3, 167

**Effective Concentration at** $x$% A type of CEC, the concentration which induces an effect of $x$% compared to the control experiment. 3, 167

**Global Effect Concentration at** 5% Another type of PNEC, the concentration which is intended to protect 95% of the global response of the community. 62, 168

**Hazardous Concentration for 5% of the species** A type of PNEC, the concentration which is intended to protect 95% of the species in the community. 4, 168

**Lethal Concentration for $x$% of the organisms** A type of CEC, the concentration which induces a mortality of $x$% compared to the control experiment. 78, 168

**Lowest Observed Effect Concentration** The lowest tested concentration which shows a statistically significant effect compared to the control. 39, 168

**No Effect Concentration** A concentration which has no effect on the endpoint considered. The existence of a NEC is an arbitrary assumption of threshold concentration effect models. 3, 168

**No Observed Effect Concentration** The highest tested concentration which does not show a statistically significant effect compared to the control. Extensively disparaged for resulting from an incorrect interpretation of statistical significance tests and for being strongly dependant on experimental design. 3, 168

**Predicted No Effect Concentration** Concentration estimated to have no effect on a target community. Can be estimated from any tier, SSD or other. 62, 168

**Quantile-Quantile plot** Plot of the theoretical quantiles of a fitted distribution against the observed quantiles of the sample. The fit of the distribution is assessed by comparing the points to a line representing exact match between the theoretical and observed quantiles. 12, 168

**Rapid Toxicity Testing** A method to rapidly sample a large number of species. The focus is on having the best characterisation of the variability in the community rather than a precise estimate for each species tolerance [Kefford et al., 2005a]. 31, 168

**Species Sensitivity Distribution** A commonly used method in Ecological Risk assessment (ERA). x, 169

# Acronyms

**ABC** Approximate Bayesian Computation. 106, 167, *Glossary:* Approximate Bayesian Computation

**ACR** Acute to Chronic Ratio. 100, 167, *Glossary:* Acute to Chronic Ratio

**AIC** Akaike Information Criterion. 13, 20, 51, 52, 141, 165, 167, *Glossary:* Akaike Information Criterion

**BC** Brain-Cousens. 47, 48, 50, 52

**CDF** Cumulative Distribution Function. 7, 14, 16–18, 26–28, 31, 167, *Glossary:* Cumulative Distribution Function

**CEC** Critical Effect Concentration. 3, 4, 7, 13, 19, 21, 22, 25–27, 29, 38, 39, 42, 54, 63, 65, 69, 70, 73, 74, 97, 100, 105, 130, 165–167, *Glossary:* Critical Effect Concentration

**COP** Conferences Of the Parties. ix

**CRS** Cedergreen-Ritz-Streibig. 47–50, 52

**DDT** dichlorodiphenyltrichloroethane. 12

**DEB** Dynamic Energy Budget. 100

**DEBtox** Dynamic Energy Budget ecotoxicological model. 75, 100

**DIC** Deviance Information Criterion. 20, 165, 167, *Glossary:* Deviance Information Criterion

$E_TX$ $E_TX$. 167, *Glossary:* $E_TX$

$EC_{50}$ Effect Concentration at 50%. 3, 33–35, 38, 42, 43, 47, 48, 52–54, 59, 62–65, 70, 71, 74, 167, *Glossary:* Effect Concentration at 50%

$EC_x$ Effective Concentration at $x$%. 3, 21, 43, 46, 47, 53, 63, 65, 66, 69, 75, 91, 96, 167, *Glossary:* Effective Concentration at $x$%

**ECHA** European Chemical Agency. 8, 19, 26, 33, 35, 100

**ERA** Ecological Risk assessment. 166

**EU** European Union. 8

$GEC_5$ Global Effect Concentration at 5%. 62–64, 105, 107, 168, *Glossary:* Global Effect Concentration at 5%

**GUTS** General Unified Threshold model for Survival. 78–82, 160

$HC_5$ Hazardous Concentration for 5% of the species. 4, 8, 10, 12–15, 17, 18, 20–22, 29, 30, 33–38, 43, 53, 62–66, 69–71, 74, 75, 78, 95–97, 100, 102, 103, 105–107, 131, 165, 168, *Glossary:* Hazardous Concentration for 5% of the species

$HC_p$ Hazardous Concentration for $p$% of the species. 31, 42, 43, 72, 74

**IT** Individual Threshold. 78, 81

$LC_{50}$ Lethal Concentration for 50% of the organisms. 4, 6, 7, 31, 33, 76, 95–97, 100

$LC_x$ Lethal Concentration for $x$% of the organisms. 78, 95, 168, *Glossary:* Lethal Concentration for $x$% of the organisms

**LKJ** Lewandowski-Kurowicka-Joe [Lewandowski et al., 2009]. 88, 89

**LOEC** Lowest Observed Effect Concentration. 39, 168, *Glossary:* Lowest Observed Effect Concentration

**NEC** No Effect Concentration. 3, 43, 75, 78, 84, 89, 91, 95–98, 100, 102, 105, 168, *Glossary:* No Effect Concentration

**NOEC** No Observed Effect Concentration. 3, 13, 39, 43, 74, 75, 168, *Glossary:* No Observed Effect Concentration

**PAF** Potentially Affected Fraction. 8

**PNEC** Predicted No Effect Concentration. 62, 107, 165, 166, 168, *Glossary:* Predicted No Effect Concentration

**Q-Q plot** Quantile-Quantile plot. 12, 16, 18, 69, 168, *Glossary:* Quantile-Quantile plot

**RSE** Residual Square Error. 13

**RTT** Rapid Toxicity Testing. 31, 75, 76, 86, 97, 105, 168, *Glossary:* Rapid Toxicity Testing

**SD** Stochatic Death. 78, 81

**SSD** Species Sensitivity Distribution. x, xi, 2–4, 8–10, 12–17, 19–22, 25–29, 31, 33, 34, 37–39, 41–43, 52, 54, 59, 62–66, 69–75, 86, 91, 95, 97–101, 103, 105–107, 125, 129–131, 169, *Glossary:* Species Sensitivity Distribution

**TD** Toxico Dynamic. 75, 78, 80, 81, 97, 103, 106

**TK** Toxico Kinetic. 75, 78, 81, 97

**TKTD** Toxico-Kinetic Toxico-Dynamic. 73, 75, 76, 78, 79, 82–84, 97, 100, 101, 106

**US** United States of America. 8

## Résumé

La SSD (Species Sensitivity Distribution) est une méthode utilisée par les scientifiques et les régulateurs de tous les pays pour fixer la concentration sans danger de divers contaminants sources de stress pour l'environnement. Bien que fort répandue, cette approche souffre de diverses faiblesses sur le plan méthodologique, notamment parce qu'elle repose sur une utilisation partielle des données expérimentales. Cette thèse revisite la SSD actuelle en tentant de pallier ce défaut. Dans une première partie, nous présentons une méthodologie pour la prise en compte des données censurées dans la SSD et un outil web permettant d'appliquer cette méthode simplement. Dans une deuxième partie, nous proposons de modéliser l'ensemble de l'information présente dans les données expérimentales pour décrire la réponse d'une communauté exposée à un contaminant. A cet effet, nous développons une approche hiérarchique dans un paradigme bayésien. A partir d'un jeu de données décrivant l'effet de pesticides sur la croissance de diatomées, nous montrons l'intérêt de la méthode dans le cadre de l'appréciation des risques, de part sa prise en compte de la variabilité et de l'incertitude. Dans une troisième partie, nous proposons d'étendre cette approche hiérarchique pour la prise en compte de la dimension temporelle de la réponse. L'objectif de ce développement est d'affranchir autant que possible l'appréciation des risques de sa dépendance à la date de la dernière observation afin d'arriver à une description fine de son évolution et permettre une extrapolation. Cette approche est mise en œuvre à partir d'un modèle toxico-dynamique pour décrire des données d'effet de la salinité sur la survie d'espèces d'eau douce.



**Mots-clefs : SSD, modèle hiérarchique, modèle Toxico-Dynamique Toxico-Cinétique, salinité, pesticides, diatomées, données censurées, protection des communautés**



## Abstract

Species Sensitivity Distribution (SSD) is a method used by scientists and regulators from all over the world to determine the safe concentration for various contaminants stressing the environment. Although ubiquitous, this approach suffers from numerous methodological flaws, notably because it is based on incomplete use of experimental data. This thesis revisits classical SSD, attempting to overcome this shortcoming. First, we present a methodology to include censored data in SSD with a web-tool to apply it easily. Second, we propose to model all the information present in the experimental data to describe the response of a community exposed to a contaminant. To this aim, we develop a hierarchical model within a Bayesian framework. On a dataset describing the effect of pesticides on diatom growth, we illustrate how this method, accounting for variability as well as uncertainty, provides benefits to risk assessment. Third, we extend this hierarchical approach to include the temporal dimension of the community response. The objective of that development is to remove the dependence of risk assessment on the date of the last experimental observation in order to build a precise description of its time evolution and to extrapolate to longer times. This approach is build on a toxico-dynamic model and illustrated on a dataset describing the salinity tolerance of freshwater species.



**Keywords: SSD, hierarchical model, Toxico-Kinetic Toxico-Dynamic model, salinity, pesticides, diatoms, censored data, protection of communities**

Appendix I

# Résumé étendu

## I.1 Introduction

La SSD (Species Sensitivity Distribution) est la clef de voûte de l'analyse du risque en écotoxicologie. Elle est utilisée par les scientifiques et les régulateurs de tous les pays pour fixer la concentration sans danger de divers contaminants, qu'ils soient issus de l'industrie ou de l'activité humaine en général. Bien que fort répandue, cette approche souffre de nombreuses faiblesses sur le plan méthodologique, car elle repose sur une utilisation partielle des données expérimentales et parfois sur l'application erronée de tests statistiques. L'hypothèse de base de la SSD est que les sensibilités à un contaminant d'une communauté d'espèces peuvent être décrites par une distribution de probabilité. Les données écotoxicologiques sont vues comme un échantillon de cette distribution et sont utilisées pour estimer la SSD. A partir de ce qui est observé au niveau d'espèces prises indépendamment les unes des autres, la SSD est alors utilisée pour extrapoler un niveau de protection pour la communauté. Le niveau de protection couramment utilisé est la $HC_5$ (Hazardous Concentration for 5% of the community), la concentration qui est sans risque pour 95% de la communauté. L'objectif de cette thèse est de revisiter l'approche SSD classique et de la développer pour permettre une utilisation plus complète des données expérimentales. Dans cette optique, la thèse est construite autour de trois axes principaux :

1. Le premier axe consiste en une revisite de la SSD actuelle, qui souligne ses manques et propose de les combler. En effet, les méthodes d'ajustement sont parfois inadaptées ou utilisées à mauvais escient. Par ailleurs, lorsque certaines données sont censurées, c'est à dire qu'elle sont disponibles sous la forme d'un intervalle (pas nécessairement borné) plutôt que d'une valeur ponctuelle, elles ne sont généralement pas prises en compte dans l'analyse. On propose un méthode pour la prise en compte des données censurées et l'on présente un outil web permettant de l'utiliser de manière automatique. L'utilisateur n'a qu'à saisir ses données, procéder à quelques choix très simples et l'analyse est effectuée en ligne immédiatement.

2. Le deuxième axe tourne autour de la proposition d'une nouvelle approche de la SSD permettant de prendre en compte toutes l'information des données expérimen-

tales, contrairement à l'approche actuelle. Cette dernière résume chaque courbe concentration-effet par un unique paramètre (une NOEC ou une $EC_x$), ce qui induit une perte considérable d'information. Pour remédier à ce problème, on cherche à ne plus résumer la réponse d'une espèce à un contaminant par une seule valeur mais par l'ensemble des paramètres d'un modèle concentration-effet caractérisant l'effet du contaminant. Cela permet d'utiliser l'ensemble de l'information présente dans les données, de modéliser toute la variabilité biologique observée lors des expériences et de propager correctement l'incertitude entre les différents niveaux de l'analyse du risque. À l'aide de ce modèle et de l'information supplémentaire extraite des données, on propose un nouvel indicateur de risque environnemental qui vient compléter celui de la SSD classique. On montre également que résumer les données dans le cadre de la SSD classique peut mener à sous-estimer de manière drastique l'incertitude sur la SSD.

3. Le troisième axe propose de prendre en compte la dimension temporelle des données expérimentales. Les mesures de reproduction ou de mortalité sont souvent relevées à plusieurs dates, mais l'approche SSD actuelle n'utilise que les données en fin d'essai. Ainsi, le résultat des mesures de la sensibilité d'une espèce dépendent de contraintes expérimentales liées à la durée de l'expérience, qui *in fine* influent sur le niveau de protection déterminé pour la communauté. On peut restreindre ce côté arbitraire en prenant en compte l'aspect dynamique des données pour conserver le maximum d'information, et ainsi proposer une détermination plus robuste du niveau de protection. Cette prise en compte s'effectue à l'aide d'un modèle mécaniste de la survie intégré dans un modèle hiérarchique développé dans le même esprit que pour le deuxième axe. On développe une SSD évoluant explicitement avec le temps et tendant vers une SSD limite indépendante du temps. On observe qu'une SSD classique estimée sur les données en fin d'expérience n'est pas aussi protectrice que la SSD limite et qu'avec la méthodologie traditionnelle, il faudrait prolonger la durée des expériences pour parvenir à un niveau de protection suffisant.

## I.2 Revisite de la SSD actuelle : prise en compte des données censurées

Les données censurées sont en général exclues de la SSD ou transformées avant utilisation, ce qui induit un biais incontrôlé dans l'estimation de la $HC_5$. On montre d'abord qu'il est relativement aisé d'adapter la méthode SSD standard pour prendre en compte les données censurées, puis on présente MOSAIC_SSD, une plate-forme web intègre cette méthode et qui permet de l'appliquer à un jeu de données arbitraire à travers une interface conviviale. On justifie les choix méthodologiques qui ont présidé à l'élaboration de MOSAIC_SSD, puis la nécessité de prendre en compte les données censurées à travers deux exemples et une étude plus systématique. On montre en particulier qu'exclure les données censurées peut induire plusieurs sortes de biais dans la SSD, notamment dans l'estimation des intervalles de confiance sur les concentrations sans risque pour l'environnement. Le travail présenté ici a fait l'objet d'une publication dans *Environmental Toxicology and*

*Chemistry*[Kon Kam King et al., 2014].
Deux jeux de données sont utilisés pour illustrer l'intérêt de la prise e compte des données censurées et pour étudier l'effet de l'exclusion et de la transformation de données sur la $HC_5$ prédite. Le premier jeu de données censuré décrit la tolérance à la salinité de macro-invertébrés d'eau douce. La tolérance de chaque espèce est résumée par une $LC_{50}$, c'est à dire la concentration qui induit une mortalité de 50% des organismes par rapport au témoin. Dans ce jeux de données, parmi 108 $LC_{50}$ disponibles, 89 (82.4%) sont censurées, dont 60 (55.6%) censurées à droite et 29 (26.8%) censurées par intervalle. La plupart des données censurées concernent des espèces pour lesquelles le faible nombre d'organismes collectés n'a pas permis l'ajustement d'un modèle concentration-réponse[Kefford et al., 2006]. Le jeu de données a été constitué de manière à être aussi représentatif que possible des espèces présentes dans la nature[Kefford et al., 2006]. Un premier avantage lié à la prise en compte des données censurées consiste à éviter d'exclure ou de transformer la majorité des données. La SSD obtenue est ainsi plus représentative de la communauté que l'on cherche à protéger. Ensuite, n'utiliser que les données non censurées introduit un fort biais de sélection vers les espèces abondantes, ce qui est particulièrement problématique dans la mesure où les espèces rares pourraient justement être celles que le régulateur souhaiterait protéger. Le second jeu de données a été publié par Koyama et al. [Koyama, 1996]. Il décrit la susceptibilité à la déformation vertébrale de poissons marins exposés au trifluralin. Le trait mesuré est l'$EC_{50}$ à 96h. 4 des $EC_{50}$ sont censurées, parmi lesquelles deux sont censurées à droite et deux à gauche. Sur ce jeu de données, prendre en compte les données censurées présente l'avantage évident de pouvoir ajuster une SSD sur 10 espèces au lieu de 6 en excluant les données censurées, quand les recommandations de l'ECHA sont de minimum 10 espèces, 15 de préférence[Aldenberg and Rorije, 2013].
Une version non censurée de ces deux jeux de données a été générée en suivant la procédure habituelle d'exclure les données censurées à droite et à gauche, et de remplacer les données censurées par intervalle par le centre de l'intervalle. Ajuster une SSD log-normale sur les jeux de données originaux et sur leur version non censurée permet de révéler l'effet pernicieux de la transformation de données sur la $HC_5$(Figure I.1).
Pour le jeu de données sur la salinité, exclure les données censurées à droite induit un biais vers le haut de la courbe cumulée et donc une $HC_5$ plus faible (Figure I.1 gauche). L'estimation pour la $HC_5$ est respectivement de 9.85 $g.L^{-1}[8.38; 11.80]$ pour le jeu de données censuré et de 7.98 $g.L^{-1}[6.63; 9.93]$ pour le jeu de données transformées. Une $HC_5$ trop basse pourrait sembler une erreur bénigne, car utiliser le jeu de données transformé aurait abouti à protéger quand même l'environnement. Pourtant, cette valeur trop faible aurait pu déclencher l'utilisation de mesures de décontamination coûteuses qui aurait pu être mieux employées ailleurs. Par ailleurs, l'influence dans un sens ou dans l'autre de la prise en compte des données censurées dépend au cas par cas du jeu de données. Notamment, on observe un effet différent sur le jeu de données trifluralin (Figure I.1 droite). Ajuster une distribution log-normale donne une $HC_5$ à $2.4 \times 10^{-3} mg.L^{-1}[4.7 \times 10^{-5}; 2.6 \times 10^{-2}]$ pour le jeu de données censurées et à $1.7 \times 10^{-2} mg.L^{-1}[8.9 \times 10^{-3}; 4.3 \times 10^{-2}]$ pour le jeu de données transformé. Exclure les données censurées aboutit à sous-estimer la variabilité de la communauté échantillonnée, ce qui conduit à une sous-estimation de la toxicité réelle du trifluralin et de son impact

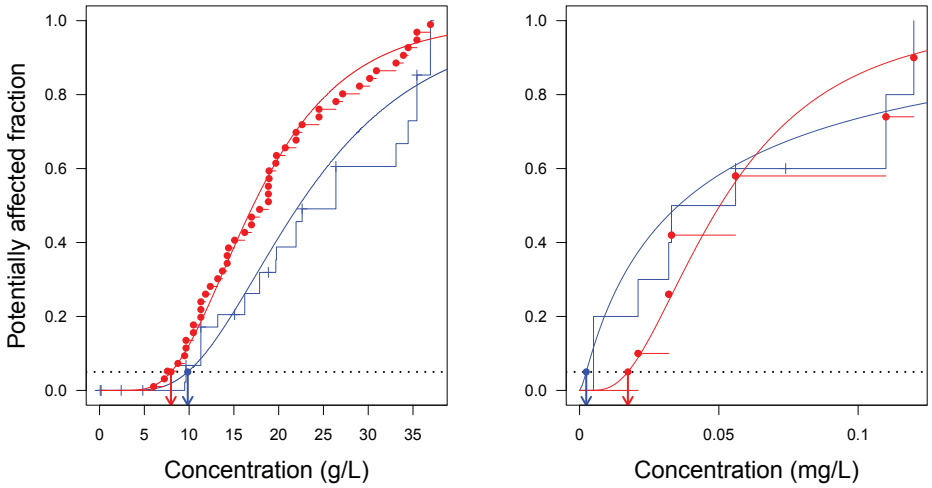



Figure I.1: Distribution cumulée ajustée et empirique avec la $HC_5$ pour le jeu de données salinité (gauche) et pour le jeu de données trifluralin (droite). La ligne pointillée correspond à une fraction potentiellement affectée de 5%. Les flèches verticales indiquent la $HC_5$. La couleur bleue indique le jeu de données censurés, la rouge les jeux de données transformé. Les distributions ajustées sur les jeux de données sont log-normales.

potentiel sur l'environnement. Une autre différence frappante est la largeur de l'intervalle de confiance sur la $HC_5$, beaucoup plus importante lorsque les données censurées sont inclues dans la SSD. Cela révèle qu'un possible effet de la transformation de jeux de données censurés est une sous-estimation sévère de l'incertitude sur la $HC_5$.
La méthode pour inclure les données censurées dans la SSD est intégrée dans un outil web: MOSAIC_SSD (`http://pbil.univ-lyon1.fr/software/mosaic/ssd/`). Cet outil reprend le modèle de la SSD classique (distribution log-normale) mais l'étend pour permettre d'intégrer les données censurées. Facile d'utilisation, il permet à n'importe quel utilisateur d'entrer ses données dans un formulaire pour obtenir le résultat (Figure I.2).

## I.3 Vers une nouvelle approche de la SSD : la SSD hiérarchique

Après avoir présenté une méthode pour prendre en compte tous les types de données y compris censurées, la deuxième étape dans la revisite de la SSD a été de revenir aux données brutes de bioessais pour tâcher d'en extraire le maximum d'information. On utilise un jeu de données de bioessais décrivant la croissance journalière d'une dizaine d'espèces de diatomées exposées à divers herbicides afin de développer une nouvelle approche de la SSD. Cette approche repose sur une modélisation hiérarchique des données écotoxicologiques, méthode innovante car elle permet de prendre en compte la totalité

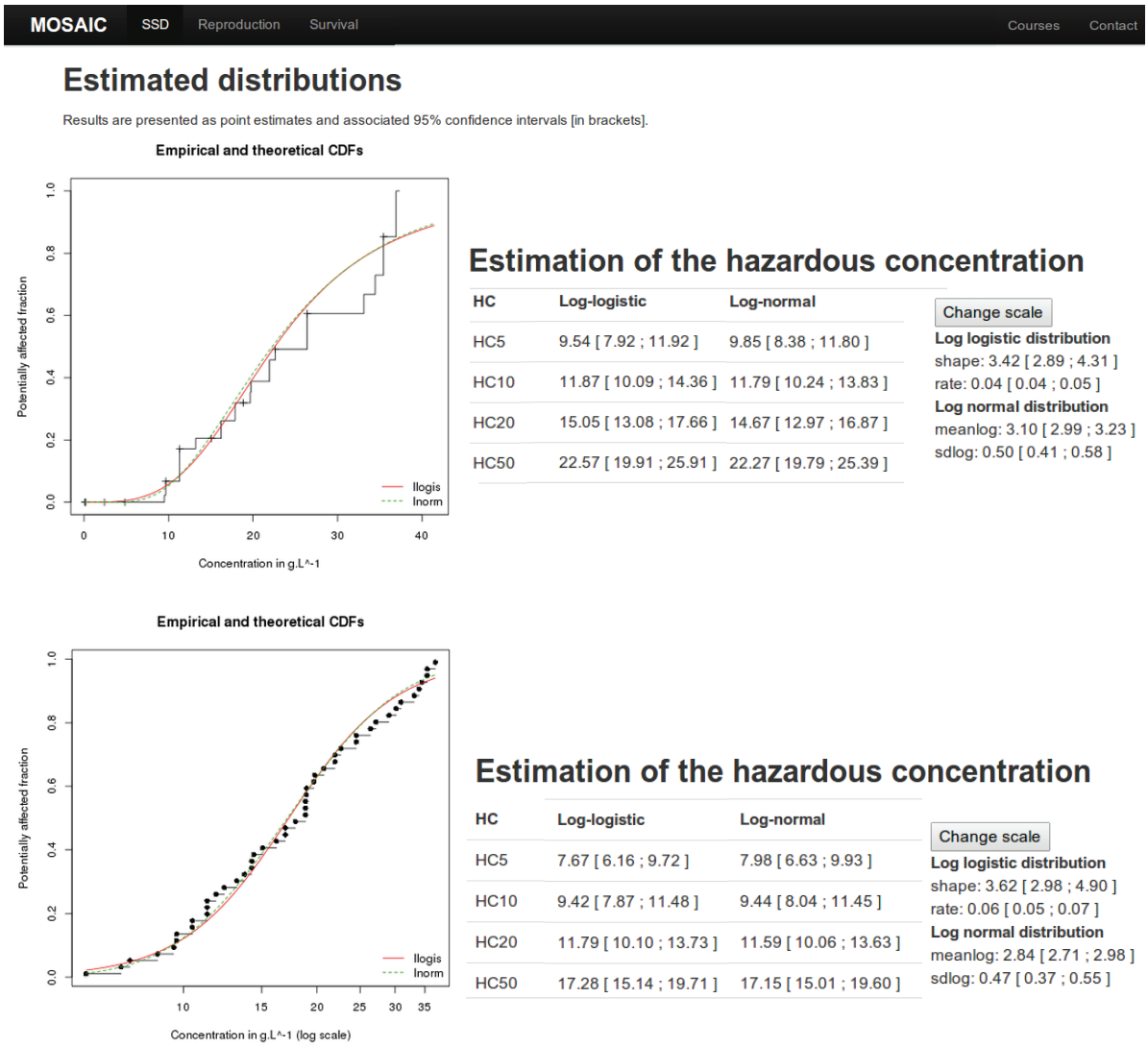



Figure I.2: Capture d'écran de la page de résultats de MOSAIC_SSD sur le jeu de données censuré de salinité. Le haut la figure montre la sortie de MOSAIC_SSD sur le jeu de données original, le bas de la figure la sortie de MOSAIC_SSD sur le jeu de données de salinity non censuré obtenu après transformation du jeu de données censuré.

.

de l'information présente dans les données expérimentales. En effet, l'approche SSD traditionnelle se contente de résumer les courbes concentration-effet par une seule valeur (e.g., la $EC_{50}$), laissant de côté des informations comme la pente de la courbe, ainsi que l'incertitude associée au modèle concentration-effet. A l'inverse, la SSD hiérarchique implique un ajustement de l'ensemble des données simultanément, ce qui permet de ne rien perdre de l'information originale. L'approche hiérarchique repose sur la modélisation de la distribution interspécifique jointe de tous les paramètres du modèle concentration-effet, et pas seulement du résumé statistique comme pour la SSD classique. Ce modèle hiérarchique a permis de proposer deux nouvelles perspectives concernant la SSD. Tout d'abord, il offre la possibilité de définir un nouvel indicateur de la réponse d'une communauté à un contaminant. La SSD classique décrit la réponse d'une communauté en terme de nombre d'espèces affectées pour une concentration donnée en contaminant. La SSD hiérarchique fournit la même information, mais elle offre également un indicateur quantitatif de la réponse globale de la communauté, liée à la variation globale de la biomasse de la communauté dans le cas précis des diatomées. Cette réponse globale est une information complémentaire à celle donnée par la SSD classique, qui ouvre la voie à une protection non plus seulement de la biodiversité de la communauté, mais aussi d'autres caractéristiques comme la biomasse, la reproduction, ou n'importe quel "endpoint" mesuré sur les espèces test. Dans le cas spécifique des diatomées, la réponse globale est comprise entre 0 et 1, et décrit la réduction relative de la biomasse globale de la communauté de diatomées pour une concentration donnée. La figure I.3 montre une comparaison entre la SSD calculée sur les $EC_{10}$, la SSD calculée sur les $EC_{50}$ et la réponse globale. En particulier, la figure permet de comparer pour deux herbicides la $HC_{5,EC_{10}}$ et la $HC_{5,EC_{50}}$ avec la $GEC_5$, c'est à dire les concentrations protégeant 95% des espèces et la concentration préservant 95% de la biomasse globale. On peut observer que dans le cas de l'atrazine, la $GEC_5$ est inférieure aux deux $HC_5$, tandis que pour le diuron, la $GEC_5$ se trouve bien en dessous de la $HC_{5,EC_{50}}$, et à peu près au niveau de la $HC_{5,EC_{10}}$. Ce dernier motif se retrouve pour les quatre autres herbicides. Il ne semble pas y avoir de relation systématique entre la $GEC_5$ et la $HC_{5,EC_{50}}$. Viser une protection de 95% de la réponse globale de la communauté peut s'avérer plus ou moins protecteur que de viser une protection de 95% des espèces (au niveau de la $EC_{10}$). Mais on peut remarquer que pour les deux herbicides, une $HC_5$ basée sur les $EC_{50}$ ne protège que 80 à 86% de la réponse globale de la communauté.

Ensuite, la SSD hiérarchique permettant un traitement rigoureux de l'incertitude associée aux données écotoxicologiques, on peut étudier l'impact de négliger cette incertitude dans l'approche SSD classique. On observe que la SSD classique induit une sous-estimation conséquente de l'incertitude associée au niveau de concentration sans risque prédit dans certains cas d'utilisation. La figure I.4 présente cette comparaison entre les prédictions de la SSD classique et celles de la SSD hiérarchique. On peut tout d'abord observer que la $HC_{5,EC_{10}}$ classique, qui néglige l'incertitude liée à l'estimation des $EC_{10}$, est bien plus haute que la $HC_{5,EC_{10}}$ hiérarchique, alors que les deux $HC_{5,EC_{50}}$ semblent coincider. On constate ensuite que l'intervalle de confiance autour de la $HC_{5,EC_x}$ hiérarchique, qui elle conserve l'incertitude originale sur les $EC_x$, s'étend rapidement pour des valeurs de x en dessous de 50. Ces différences peuvent se comprendre en partie parce que pour de faibles valeurs de x, l'incertitude sur la $EC_x$ estimée à partir du modèle concentration-

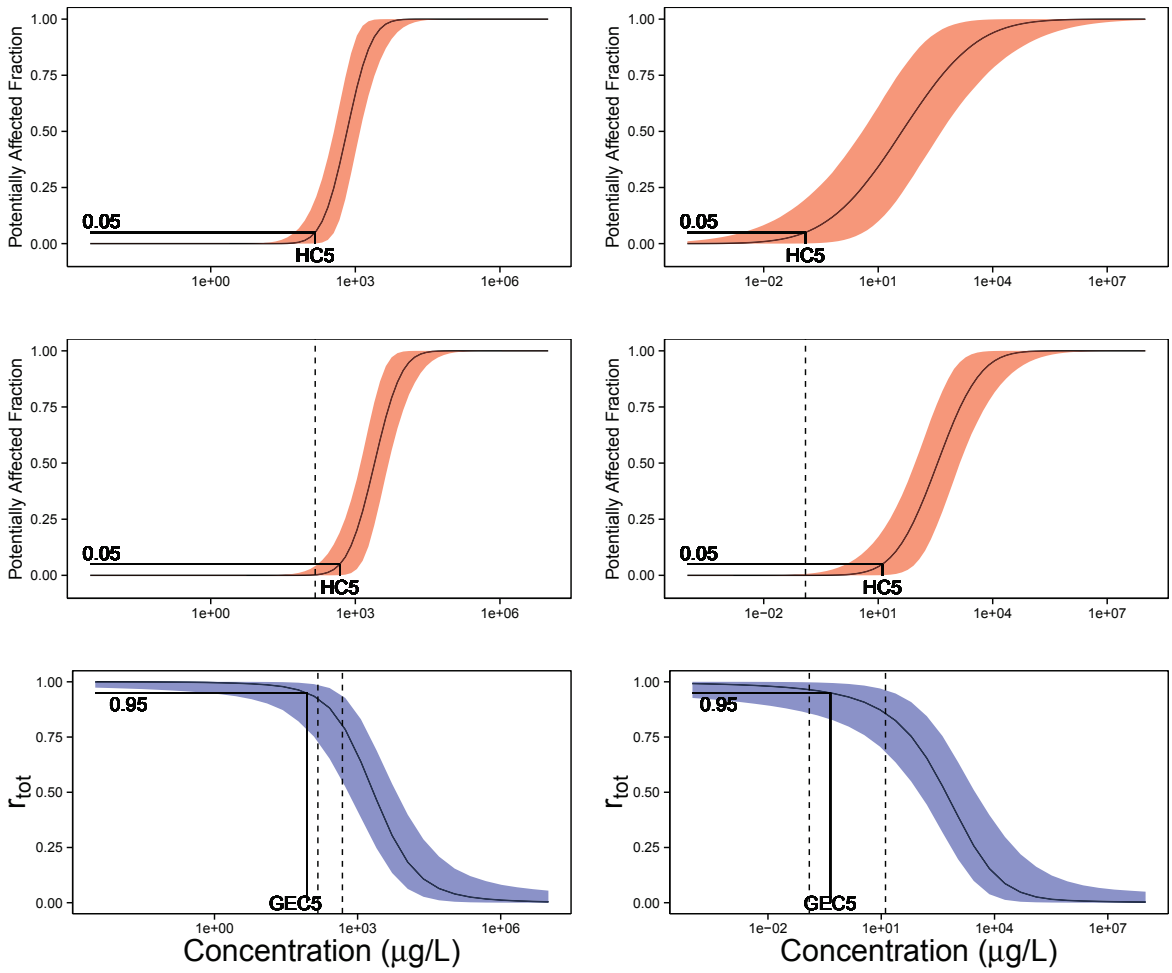



Figure I.3: SSD et réponse globale de la communauté à l'atrazine (gauche) et au diuron (droite). Haut : SSD classique, ajustée sur les $EC_{10}$ avec la bande de confiance à 95% calculée par bootstrap et la $HC_5$. Milieu : SSD classique, ajustée sur les $EC_{50}$ avec la bande de confiance à 95% calculée par bootstrap et la $HC_5$. Bas : réponse globale de la communauté avec la bande de crédibilité à 95% et la concentration correspondant à une réduction de 5% de la réponse globale ($GEC_5$). Les lignes pointillées horizontales permettent de comparer la $HC_{5,EC_{10}}$, la $HC_{5,EC_{50}}$ et la $GEC_5$

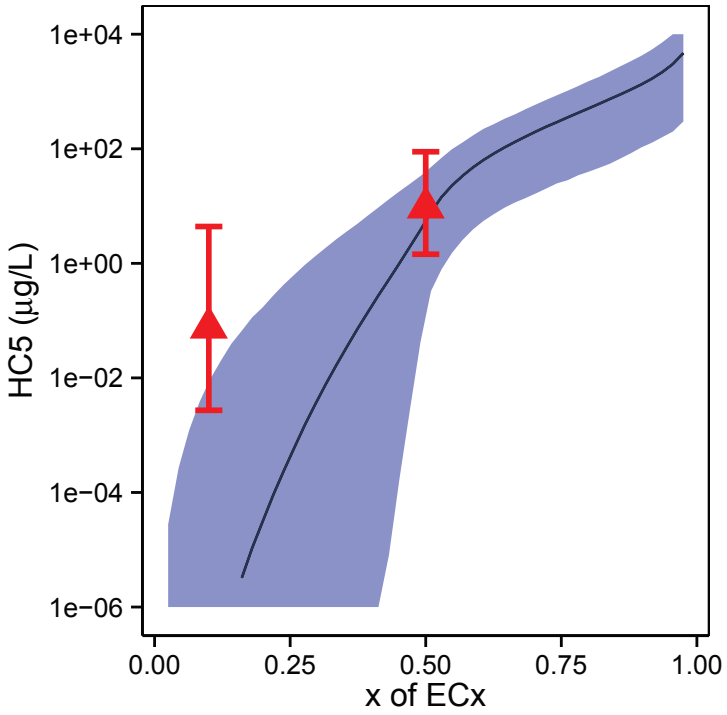



Figure I.4: $HC_5$ en fonction du $x$ de la $EC_x$ pour le diuron obtenue à partir de la SSD hiérarchique, et l'intervalle de crédibilité à 95%. En rouge figurent les $HC_5$ obtenues avec la SSD classique ajustées sur les $EC_{10}$ et les $EC_{10}$, avec un intervalle de confiance bootstrap à 95%.

effet est d'autant plus importante. Elle sont également liées à la structure du modèle hiérarchique, qui rend explicite le fait qu'une $EC_x$ est liée par une relation mathématique à la $EC_{50}$ et à la pente de la courbe concentration-effet. Si l'on suppose que la $EC_{50}$ suit une distribution log-normale, comme c'est souvent le cas dans le cadre de la SSD classique, alors une $EC_x$ n'a aucune raison de suivre la même forme de distribution.
L'un des avantages de l'approche hiérarchique réside dans la prédiction d'une réponse globale de la communauté sous la forme d'une courbe concentration-effet qui permet de déterminer une concentration d'effet globale, la $GEC_5$. Ce nouvel indicateur vient compléter la $HC_5$, plutôt que s'y substituer. En effet, la $HC_5$ vise à protéger 95% des espèces, mais il reste une incertitude considérable sur le destin de la communauté si les 5% non protégés jouent un rôle clef pour certaines propriétés de la communauté. La $GEC_5$, elle, vise à protéger 95% d'une réponse globale mais ne renseigne pas sur la proportion des espèces affectées de manière significative, donc *in fine* sur l'atteinte à la biodiversité.

## I.4 Vers une nouvelle approche de la SSD : prise en compte de la dimension temporelle

Il y a deux motivations pour prendre en compte la dimension temporelle de la SSD. Tout d'abord, dans le souci d'utiliser au mieux les données issues de bioessais, une suite logique des travaux précédents consiste à modéliser la dimension temporelle des données négligée jusqu'ici. En effet, un bioessai comporte couramment un suivi temporel, mais en général seules les données en fin d'essai sont utilisées. La deuxième motivation provient du fait que la sensibilité d'une espèce à un contaminant dépend du temps d'exposition, or pour des raisons pratiques la majorité des tests sont menés sur de courtes durées. Les SSD classiques calculées alors sont des descriptions très imparfaites de la sensibilité d'une communauté, qu'il est possible d'améliorer en tirant un meilleur parti des données déjà collectées.
Pour intégrer cette dimension temporelle à la SSD dans le cadre de mesures de survie, l'approche générale retenue comprend une modélisation mécaniste de la survie en fonction du temps, puis de l'influence de la concentration en contaminant sur le taux de survie à l'aide d'un modèle à seuil de type NEC (No Effect Concentration). Ensuite, on construit un modèle hiérarchique de la SSD dans le même esprit que pour la section précédente, que l'on ajuste sur les données. L'utilisation d'un modèle de type NEC est particulièrement intéressante pour la SSD, parce que la sensibilité des espèces est définie par une concentration sans effet. Dès lors, une $HC_5$ construite sur les NEC et qui protège 95% des espèces signifie bien qu'à une concentration égale à la $HC_5$, 95% des espèces ne subissent pas d'effet du polluant. Pour une $HC_5$ construite sur des $EC_{50}$, la $HC_5$ est une concentration pour laquelle 95% des espèces sont affectées à moins de 50% d'effet, ce qui peut signifier 1% comme 49%. Or selon le cas, il peut être problématique de considérer des espèces affectées à 49% comme étant protégées.

### I.4.1 Description des données

Dans le cadre d'une collaboration avec Ben Kefford et Christophe Piscart, j'ai eu accès à des données de salinité proviennent d'expériences de survie menées sur des espèces d'eau douce collectées en France et en Australie. Ce jeu de données se distingue par sa taille et son hétérogénéité : il contient 217 espèces pour lesquelles le nombre de concentrations mesurées varie de une à plusieurs dizaines. Le nombre d'individus survivants en fonction de la concentration a été relevé tous les jours pendant 3 jours pour toutes les espèces, et un jour supplémentaire pour une partie des espèces. Les concentrations sont données en $\mu S.cm^{-1}$, unité de conductivité ionique. Il est à noter que la salinité a une toxicité chronique supérieure à la toxicité mesurée sur 3 ou 4 jours, donc les données en fin d'expérience ne reflètent pas la toxicité réelle de la salinité sur une longue durée. Les espèces ont été collectées au hasard et à grande échelle pour couvrir autant de groupes taxonomiques que possible, et pour disposer d'un échantillon représentatif des espèces présentes dans les régions de collecte. Les espèces peuvent être séparées entre espèces *rares* et *communes*. Alors qu'il n'y a que quelques mesures de concentration pour les premières, il peut y avoir de nombreuses concentrations et même des réplicats pour les secondes. Une précédente étude de ce jeu de données a conclu a une relative simili-

tude entre les jeux de données provenant de France et d'Australie[Kefford et al., 2012a], justifiant un rassemblement de toutes les espèces en un seul jeu de données. Plus précisément, il a été prouvé que bien que les espèces françaises semblent un peu plus sensibles, la présence de divers groupes taxonomiques est une source de variabilité bien plus importante que la région de collecte.

## I.4.2 Modèle TKTD (Toxico-Kinetic Toxico-Dynamic)

### I.4.2.1 Introduction

Les modèles TKTD sont une classe de modèles *mécanistes* de l'effet d'un contaminant sur la survie, au sens où ils intègrent une description explicite du mécanisme menant à la mort de l'individu. Ce processus est décrit en deux phases, une phase toxicocinétique qui modélise l'entrée progressive du contaminant dans l'organisme, et une phase toxicodynamique qui modélise les dégâts graduels causés par le contaminant au sein de l'organisme et menant à sa mort. Les avantages d'un modèle mécaniste résident dans le (relativement) faible nombre de paramètres, et dans le sens biologique qu'on peut leur attribuer, contrairement à un modèle phénoménologique.
Le modèle GUTS[Jager et al., 2011] (General Unified Threshold model of Survival) appartient à cette classe de modèles TKTD mécanistes, avec la particularité que la partie toxicodynamique comprend un modèle à seuil, c'est à dire un paramètre délimitant une concentration en dessous de laquelle le contaminant n'a pas d'effet sur la survie de l'individu.

### I.4.2.2 Modèle

Les modèles à seuil, très courants en toxicologie font généralement l'hypothèse que le seuil est indépendant du temps et n'est déterminé que par un seul paramètre. Cette hypothèse a été conservée. Ces modèles à seuil n'incluent en général pas de phénomène d'hormèse ou d'essentialité. Le modèle utilisé ici suppose une décroissance monotone de la probabilité de survie avec le temps et la concentration. Le modèle GUTS a été adapté pour l'étude de la toxicité de la salinité.
La probabilité de survie à un instant $t$ est définie comme la probabilité que la mort de l'individu survienne au delà de cet instant.

$$S(t) = P(T > t) \tag{I.1}$$

On définit $h_z(t)$ (*hazard rate*) le taux de mortalité par la relation suivante :

$$h_z(t) = -\frac{1}{S(t)}\frac{dS(t)}{dt} \tag{I.2}$$

Le modèle à seuil est défini, pour une exposition à concentration constante, sur ce taux de mortalité :

$$h_z = m^0 + k^s \left(C^w \left(1 - e^{-k^r t}\right) - NEC\right)_+ \tag{I.3}$$

où $C^w$ est la concentration dans l'eau, $m^0$ est la mortalité naturelle, et $k^s$ est un paramètre en $jours^{-1}concentration^{-1}$ qui contrôle la dépendance du taux de mortalité à la con-

centration. $k^r$ est un paramètre en $jours^{-1}$ qui contrôle le délai entre l'exposition au contaminant et l'effet sur la survie. $(x - NEC)_+$ est une fonction de Heaviside qui vaut 0 si $x$ est inférieur à la $NEC$, et qui vaut $x - NEC$ sinon. Enfin, le paramètre $NEC$ est la *No Effect Concentration*, la concentration en dessous de laquelle le contaminant n'a aucun effet. Cette absence d'effet est implémentée mathématiquement par la fonction de Heaviside.

Pour $C^w > NEC$, il y a un temps $t_{NEC}$ avant lequel un organisme avec un niveau de dommage initial négligeable ne ressentira pas l'effet du contaminant sur sa survie :

$$t_{NEC} = -\frac{1}{k^r} \ln\left(1 - \frac{NEC}{C^w}\right) \tag{I.4}$$

Pour $t < t_{NEC}$:

$$S(t) = e^{-m^0 t} \tag{I.5}$$

et pour $t > t_{NEC}$

$$S(t) = e^{-\int_0^t h(u)du} \tag{I.6}$$

$$= e^{-m^0 t - k^s\left[(C^w - NEC)(t - t_{NEC}) + \frac{C^w}{k^r}\left(e^{-k^r t} - e^{-k^r t_{NEC}}\right)\right]} \tag{I.7}$$

Enfin, afin d'ajuster le modèle aux données, on a besoin de la fonction de vraisemblance des données sachant les paramètres du modèle. On utilise un modèle d'erreur binomial pour modéliser la survie d'un individu à un temps donné, sachant qu'il était vivant à un temps précédent. On écrit la probabilité conditionnelle de survie de la manière suivante :

$$P(T > t'|T > t) = \frac{P(T > t')}{P(T > t)} = \frac{S(t')}{S(t)} = e^{-(m_0 + k_s \mathcal{H}_{NEC}(x)(x - NEC))(t' - t)} \tag{I.8}$$

pour $t' > t$.

La fonction de vraisemblance pour les données d'une espèce se déduit de la relation :

$$N_{x,t'} \sim \mathcal{B}(N_{x,t}, P(T > t'|T > t)) \tag{I.9}$$

où $\mathcal{B}$ représente la distribution binomiale, et $N_{x,t}$ le nombre de survivants à la concentration $x$ et au temps $t$.

## I.4.3 Modèle hiérarchique

### I.4.3.1 Présentation du modèle hiérarchique

Le modèle TKTD représente l'étage inférieur du modèle hiérarchique. C'est l'étage monospécifique, qui décrit la réponse de chaque espèce. À l'étage supérieur se trouve le modèle décrivant la distribution des paramètres du modèle TKTD au sein de la communauté. Il est construit selon une stricte analogie avec l'approche SSD classique. En effet,

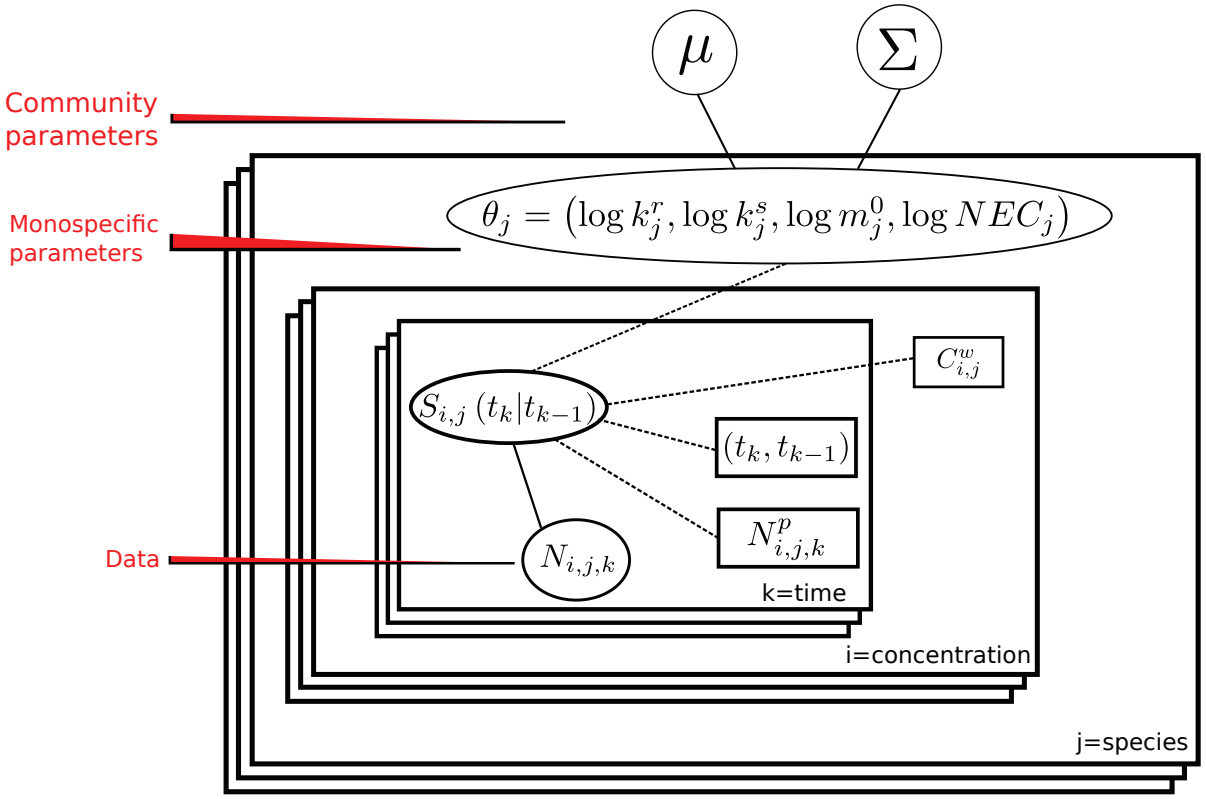



Figure I.5: Diagramme Acyclique Dirigé (DAG) du modèle hiérarchique. Les flèches en pointillé représentent un lien stochastique, celles en traits pleins un lien déterministe.

dans le cadre de la SSD classique, la sensibilité de chaque espèce est caractérisée par un paramètre, la $LC_{50}$, et on suppose que la distribution des $LC_{50}$ dans la communauté suit une loi log-normale. Dans le cadre de la SSD hiérarchique, la sensibilité d'une espèce est caractérisée par quatre paramètres, $k^r$, $k^s$, $NEC$ et $m^0$ . On suppose que la distribution de ces quatre paramètres dans la communauté est également log-normale. Il est naturel de considérer une éventuelle corrélation entre les paramètres, donc on choisit une distribution log-normale multivariée qui prend en compte ces corrélations via sa matrice de variance-covariance. Le diagramme acyclique dirigé (DAG) représenté sur la figure I.5 donne une représentation graphique du modèle hiérarchique, qui met en évidence les différents niveaux du modèle. L'intérêt principal des DAG est qu'ils se traduisent presque directement sous la forme d'un modèle analysable numériquement, mais celui-ci permet déjà de visualiser quels paramètres caractérisent la distribution de sensibilité dans la communauté, quels paramètres représentent la sensibilité d'une espèce, et quelles sont les relations entre ces paramètres.

### I.4.3.2 Intérêt du modèle hiérarchique

Le modèle hiérarchique est intéressant à plusieurs égards dans le cadre de l'analyse des données de salinité. Il s'accommode bien de la grande hétérogénéité des données, et permet par ailleurs de prendre en compte des espèces sur lesquelles on ne dispose que de quelques points de mesure, c'est à dire les espèces rares qui sont nombreuses

Table I.1: Description des liens décrits dans le DAG(figure I.5)

| Noeud | Type | Équation |
|---|---|---|
| $ks_j$ | Stochastique | $k_{s_j} \sim \mathcal{N}(\mu_{ks}, \sigma_{ks})$ |
| $NEC_j$ | Stochastique | $NEC_j \sim \mathcal{N}(\mu_{NEC}, \sigma_{NEC})$ |
| $S_{i,j,t}$ | Déterministe | $S_{i,j,t} = e^{-\left(m_{0,j}+k_{sj}\mathcal{H}_{NEC_j}(c_{i,j,t})(c_{i,j,t}-NEC_j)\right)t}$ |
| $yi, j, t$ | Stochastique | $y_{i,j,t} \sim \mathcal{B}(y_{i,j,t-1}, \frac{S_{i,j,t}}{S_{i,j,t-1}})$ |

dans ce jeu de données. La structure hiérarchique implique que chaque espèce participe à la détermination de la distribution de sensibilité de la communauté à la mesure de l'information disponible sur cette espèce. Les espèces communes et bien documentées fournissent beaucoup d'information mais sont peu nombreuses, tandis que les espèces rares apportent peu d'information individuellement mais représentent la majorité des espèces.

### I.4.4 Comparaison SSD hiérarchique sur la NEC et SSD classique

On souhaite commencer par comparer l'approche NEC hiérarchique avec l'approche SSD classique. On ne détaillera pas beaucoup l'approche SSD classique ici, mais celle retenue consiste en l'ajustement d'un modèle log-logistique à deux paramètres sur la proportion de survivants pour estimer les $LC_{50}$. Quand les données ne permettent pas l'ajustement d'un modèle, la $LC_{50}$ est déterminée visuellement, sous la forme d'un intervalle qui contient vraisemblablement la $LC_{50}$. Une grande partie des $LC_{50}$ est donc censurée, et on utilise l'approche MOSAIC_SSD [Kon Kam King et al., 2014] pour ajuster une SSD log-normale.
La SSD log-normale classique est ajustée sur les $LC_{50}$ à 24, 48 and 72 heures pour obtenir une $HC_5$ à 24, 48 and 72 heures. La SSD temporelle permet de calculer la distribution des $LC_{50}$ à chaque temps, c'est à dire un équivalent de la SSD classique à chaque temps, et ses prédictions peuvent être comparées à celles de la SSD classique. La SSD temporelle permet aussi de calculer la distribution des NECs qui elles ne dépendent pas du temps. La $HC_5$ de la SSD classique et celle de la SSD temporelle, toutes deux fondées sur les $LC_{50}$ décroissent avec le temps, alors que la $HC_5$ fondée sur les NECs est invariante avec le temps (Figure I.6). La $HC_5$ fondée sur les NEC $(4407[3310 - 5634]\mu S/cm)$ est plus faible que la $HC_5$ classique à 24, 48 and 72 H (10242[8106 − 12805], 8821[7112 − 10846] et $6800[5405 - 8643]\mu S/cm$, respectivement). Prendre des décisions de gestion de l'environnement en se fondant sur la $HC_5$ calculée à partir de données de fin d'expérience (72 heures d'exposition) ou bien en se fondant sur la $HC_5$ calculée sur les NECs aurait des conséquences pratiques très différentes.
La $HC_5$ temporelle calculée sur les $LC_{50}$ au cours du temps est très similaire à la $HC_5$ classique : l'approche hiérarchique semble reproduire les résultats de l'approche classique jusqu'à 72h. Pour toutes les valeurs de $x$, la $LC_x$ converge avec le temps vers la NEC[Smit and Ebbens, 2008], ce qui implique que la $HC_5$ fondée sur les $LC_{50}$ converge aussi vers la $HC_5$ fondée sur les NEC(Figure I.6).

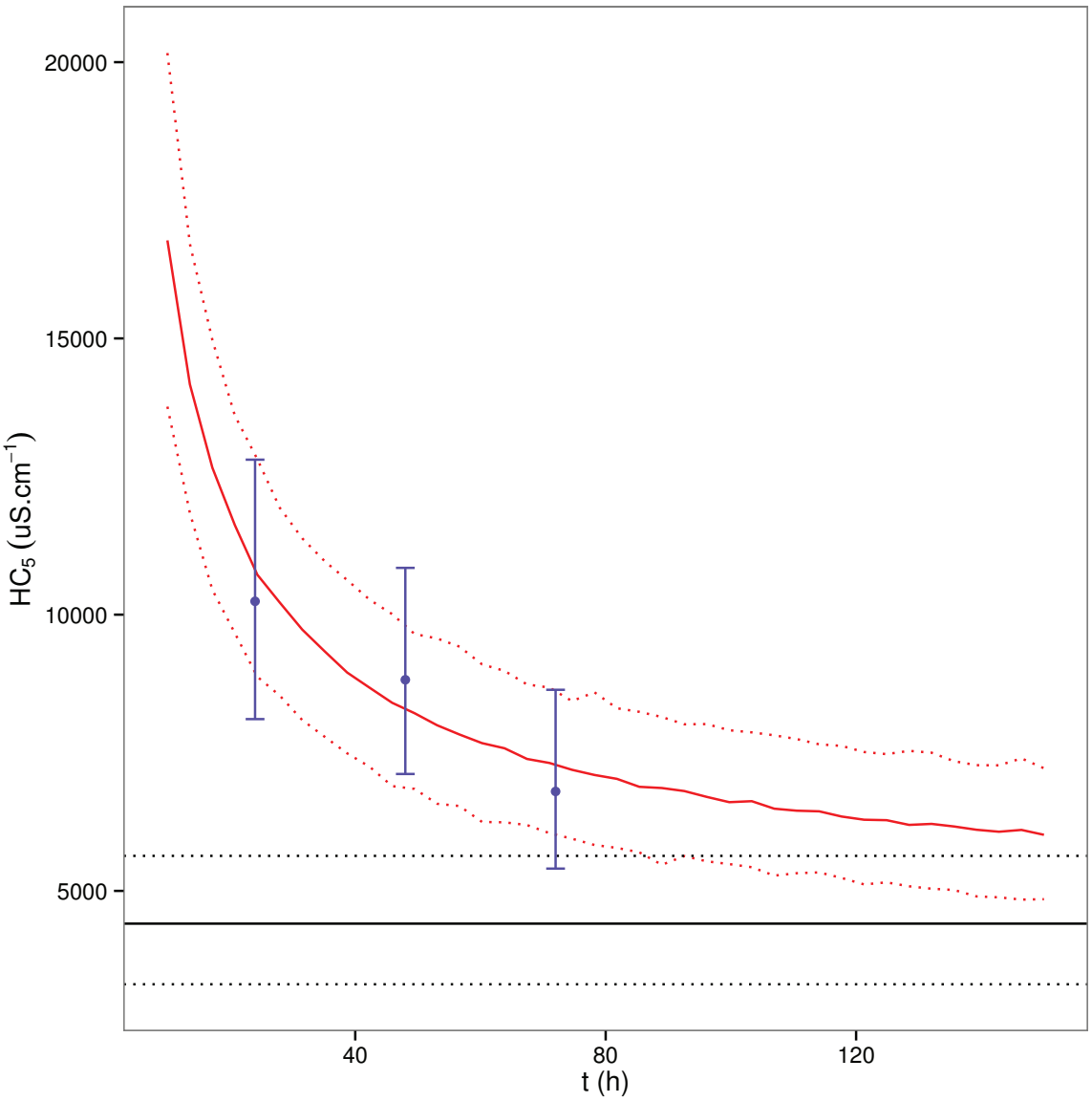



Figure I.6: $HC_5$ classique (bleu) et hiérarchique (rouge) calculées sur les $LC_{50}$ en fonction du temps et $HC_5$ hiérarchique calculée sur les NEC (noir). Les segments verticaux délimitent l'intervalle de confiance à 95% sur la $HC_5$ classique. Les lignes pointillées délimitent l'intervalle de crédibilité à 95% sur la $HC_5$ hiérarchique, les lignes pleines représentent la médiane de la distribution postérieure.

### I.4.5 Utilisation du modèle en prédiction

Le modèle dynamique de survie utilisé dans cette partie n'étant pas restreint à des scénarios d'exposition constante, il est possible de prédire la réponse d'une communauté d'espèces soumises à des pics de concentration ou n'importe quel scénario plus réaliste. A partir de suivi de données de salinité dans le bassin hydrographique du Murray Darling, en Australie, on observe l'effet d'une exposition à des fluctuations de salinité.
On compare la réponse de la communauté à deux scénarios d'exposition, l'un correspondant à ce qui a été observé et l'autre correspondant au résultat d'une possible opération de réduction de la salinité. Les deux scénarios d'exposition ont un effet très différent sur la survie (Figure I.7). Alors que la survie décroît rapidement dans le cas sans intervention, lorsque la salinité est réduite de seulement 30% on observe déjà un effet important. On pourrait comparer l'effet de mesures plus réalistes sur la survie dans la communauté en utilisant la même approche.

## I.5 Conclusion et perspectives

Cette thèse propose diverses pistes pour améliorer la SSD actuelle. Dans la première partie de la thèse, on a expliqué comment les tests de toxicité pouvaient produire des données censurées et on a présenté une méthode simple pour inclure ce type de données dans la SSD. Dans la seconde partie de la thèse, on a expliqué que la SSD classique n'utilisait qu'un résumé des données disponibles, ce qui avait pour conséquence d'oublier l'incertitude attachée au résumé en question ainsi qu'une partie de l'information biologique disponible. On a montré comment étendre la SSD pour prendre en compte cette incertitude et toute l'information disponible via un modèle hiérarchique de la courbe dose-réponse dans son ensemble. Dans la troisième partie de la thèse, on a expliqué que les bioessais était souvent réalisés avec un suivi temporel, mais qu'en général seules les données de fin d'essai étaient utilisées dans la SSD classique. Cela implique que la $HC_5$ prédite dépend de la durée de l'expérience, ce qui réduit sa validité pour la protection de l'environnement. Étendant le travail réalisé dans la deuxième partie, on a proposé d'utiliser un modèle dynamique pour modéliser des données de survie suivies au cours du temps. Cela a permis d'estimer une $HC_5$ valable quelle que soit la durée d'exposition.
Tout au long de cette thèse, on a pris soin de ne proposer que des développements de la SSD qui ne requéraient pas la collecte de données supplémentaires, mais qui se fondaient au contraire sur une meilleure utilisation des données déjà disponibles. Or, quelques récents travaux sur la SSD se sont concentrés sur l'inclusion d'information de nature différente, comme la sensibilité à d'autres contaminants, les données issues d'expériences antérieures, la sensibilité d'espèces taxonomiquement voisines ou les avis d'experts[Craig et al., 2012, O'Hagan et al., 2005]. La piste la plus intéressante d'extension du travail présenté dans cette thèse serait donc de relier ces approches pour tirer le meilleur parti des différentes sources d'information disponibles sur la sensibilité des espèces.

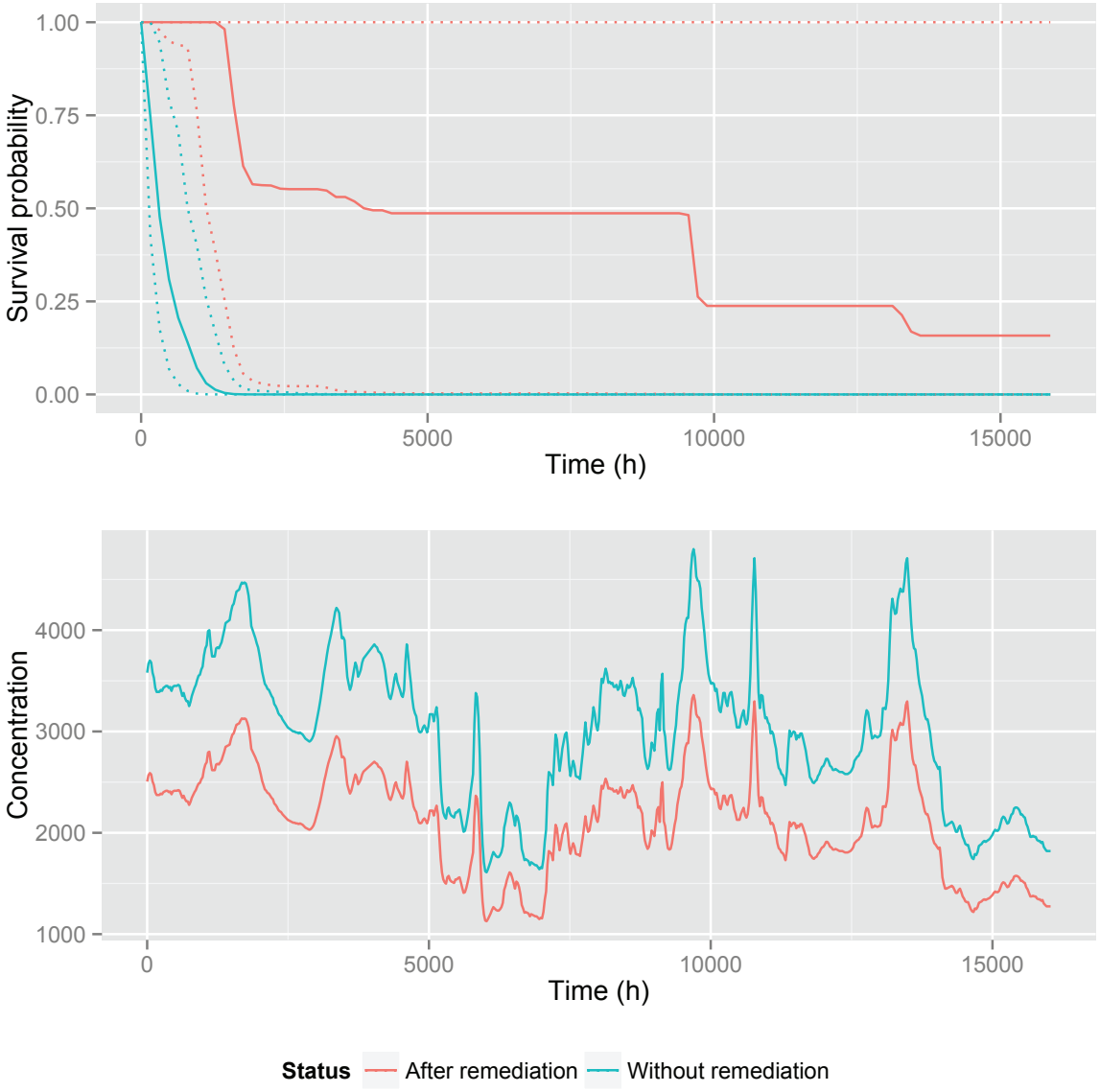



Figure I.7: Haut: Comparaison de la survie dans la communauté en fonction du temps, avec et sans l'effet du mesure de désalinisation schématique induisant une réduction de 30 % de la salinité. Les lignes pontillées représentent l'intervalle de crédibilité á 95%. Bas: profil de salinité avec et sans la mesure de désalinisation.